UNIVERSAL
LIBRARY

OU_160169

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 153.5 J66P Accession No. 34055

Author Johnson, Donald M.

Title Psychology of Thought and Judgment 1955

This book should be returned on or before the date last marked below.

The Psychology of Thought and Judgment

Under the Editorship of

GARDNER MURPHY

The Psychology of Thought and Judgment

DONALD M. JOHNSON
Michigan State College

Harper & Brothers • Publishers • New York

THE PSYCHOLOGY OF THOUGHT AND JUDGMENT

Copyright, 1955, by Harper & Brothers ©
Printed in the United States of America

All rights in this book are reserved.
No part of the book may be used or reproduced in any manner whatsoever without written permission except in the case of brief quotations embodied in critical articles and reviews. For information address Harper & Brothers
49 East 33rd Street, New York 16, N. Y.

C-E

Library of Congress catalog card number: 55–6350

Contents

PREFACE ix

1. INTRODUCTION 1

Ideas in Consciousness—Gestalt Psychology and Cognitive Patterns—Thought as an Instrument of Adjustment—Psychophysics: From Sensation to Judgment—The Testing Movement—The Nature of the Topic—References

2. THE THOUGHT PROCESSES: METHODS FOR ANALYSIS AND CLASSIFICATION 21

Subjective Procedures—Objective Experimentation—Analysis of Individual Differences—Trifunctional Analysis of Thought Processes—References

3. THE DYNAMICS OF THOUGHT 56

Motives for Thought—What Is a Problem?—Set for Single Stimulus and Response—Attention Span—Set for Continuous Intellectual Work—Individual Differences—References

4. PERCEPTION AND THOUGHT 102

Influence of Motives and Sets—Perceptual Organization and Reorganization—Individual Differences—References

5. LEARNING AND THINKING 115

Logical Analysis—Analysis of the Acquisition Phase—The Contribution of Previous Learning to Thought and Judgment—Individual Differences in Learning—References

6. PREPARATION FOR THOUGHT AND JUDGMENT 146

The Discrimination Experiment—The Association Experiment—Laboratory and Classroom Problems—Creative Thought—Individual Differences in Preparation—References

7. THE PRODUCTION PROCESS: SIMPLE RESPONSES 178
Production of Single Words—Continuous Production of Words—Production of Solutions to Simple Problems—Individual Differences in Production of Simple Responses—References

8. THE PRODUCTION PROCESS: COMPLEX PATTERNS AND PROBLEMS 221
The Functions of Patterns in Problem Solving—Describing Patterns and Their Use—How Patterns Are Formed and Used in Solving Problems—Creative Thought—Individual Differences—References

9. JUDGMENT: STIMULUS VARIABLES 280
The Function of Judgment—Judgments of Simple Material—Judgments of Complex Material—Organizing the Stimulus Material—Individual Differences—References

10. JUDGMENT: RESPONSE VARIABLES 324
Expressing the Judgment—The Response Scale—Organization of the Response Scale—Confidence, Time, and Difficulty—Individual Differences—References

11. INDIVIDUAL DIFFERENCES IN THOUGHT AND JUDGMENT 390
General Problem-Solving Ability—General Intelligence as Problem-Solving Ability—Group Factors and Problem-Solving Abilities—Reasoning—Good Judgment as an Ability—Creative Thinking—Implications for the Analysis of Thinking—References

12. INTELLECTUAL ACTIVITIES OF SOCIAL GROUPS 445
Methods of Investigation—Some Determinants of Group Problem Solving—Some Determinants of Group Judgments—Are Groups Superior to Individuals?—Social Organization of Problem Solving—References

13. THE IMPROVEMENT OF THINKING 481
General Recommendations—Preparation—Improvement of

Production—Improvement of Judgment—Improvement of Group Thinking—References

INDEX OF NAMES 503

INDEX OF SUBJECTS 511

Preface

A VOLUME like this, which attempts to systematize an ill-defined field of investigation, could be written in several ways. The author has chosen to describe and organize what is known about each section of the field and to push a short distance into the unknown. Thus the treatment will vary from one chapter to another in several respects. On some pages the discussion will be quite rigorous; on others hardly more than common sense. Some topics will be of current interest; a few may sound old-fashioned. Some questions can be fairly well answered; others cannot even be stated with proper clarity. Throughout, the central theme is the activities of the socialized human being solving his problems. The variations are due to the uneven light that psychological progress has thrown on these activities.

Occasionally a belief is expressed that the study of thought has gone out of date. To assemble the facts on this point a tally was made of the dates of the references at the end of each chapter. The tabulation shows that the great bulk of the work cited has been published in recent years. Since 1950, there have been 185; from 1940 to 1949, there were 305; from 1930 to 1939, 172; from 1920 to 1929, 57; from 1910 to 1919, 19; from 1900 to 1909, 6; and from 1890 to 1899, 2.

People think about all kinds of things under all kinds of conditions. Hence, good data on thinking come from many sources. One aim of this book is to show the reader the variety of kinds of evidence pertinent to our topic. As to completeness: when a specific question has been studied by only a few investigators, usually all are mentioned, especially if the results are divergent. When there are many similar reports, a choice is made in favor of the earliest report or the one that can be presented most clearly to the reader. The author and, presumably, the reader are indebted to the hundreds of investigators who have struggled to get some tangible evidence on these subtle intellectual operations. We are indebted also to the publishers who have kindly given permission to use copyrighted materials.

Several omissions will be noted, such as infrahuman problem solv-

ing, autistic thought and fantasy, and developmental trends, all of which are treated elsewhere. The customary distinction between logic and the psychology of thought, e.g., between the logic of scaling and the description of scales of judgment, has been observed. There is also a vast speculative and inspirational literature which has been passed over in favor of empirical data.

A strenuous effort has been made to keep the book short so that it will serve as a convenient organization of present knowledge and guide to future research.

D. M. J.

East Lansing, Michigan
January, 1955

The Psychology of Thought and Judgment

Chapter 1

Introduction

> Tell me, O tell, what kind of thing is wit,
> Thou who master art of it.
>
> —ABRAHAM COWLEY

TO anyone with a practical or theoretical interest in human beings the status of the psychology of thought is distressing. Such fields of psychology as perception, learning, and personality are being explored thoroughly and systematically. The science of psychology may be proud of current progress in these fields. But the scientific investigation of thinking is unsystematic and, in general, unsatisfactory.

Many psychologists express an interest in thinking, and some make raids on it. The field is littered with disabled wisdoms cast off by hit-and-run theorists. The purpose of the present work is to tidy up and remap this field, charting what is known and pushing a step or two ahead.

The psychology of thought began in different ways at different times. The paths of influence leading from the past to the present crisscross through several centuries. They are clearer to us today, at this end, than they were to the pioneers who trod them. A brief sketch of the main roads to the present, as seen by a backward-looking eye, will aid in showing where the psychology of thought is today and in accounting for its failures and achievements.[1]

[1] Much of the material in this chapter is treated more thoroughly in such standard references as Boring (*1950*) and Murphy (*1949*). (Complete bibliographical information for the italic numbers in parentheses is to be found in the References at the end of each chapter.)

IDEAS IN CONSCIOUSNESS

As soon as man became self-conscious, he began to examine his consciousness. One early result of this self-observation was the separation of the mind or consciousness from the body. The psychologies of antiquity, at least those of most influence on later psychologies, were dualistic. By the beginning of the eighteenth century many educated thinkers of our civilization had become accustomed to analyzing their consciousness into ideas. Ideas were the elements of thought, which can be combined and rearranged so as to make up the complex thoughts of complex thinkers. Many convincing principles of the association of ideas were worked out by the psychological philosophers of the eighteenth century, chiefly in Great Britain.

INTROSPECTION

The British associationists were philosophers before they were psychologists and their method was logical before it was observational. Careful observation of one's own consciousness was developed in Germany in the nineteenth century and polished brightly by Wundt and his followers. One of his most influential disciples, E. B. Titchener, brought the Wundtian tradition to America. The Wundtians aimed to make psychology, defined as analysis of the contents of the mind, a precise science by refining the method of self-observation, or introspection.

Classical introspection, as developed by Wundt, was a special kind of self-examination, which required knowing what to look for and what to ignore. The elements of the mind were sensations, images, and feelings, and training in introspection consisted of practice in identifying these elements. The method was experimental in that the conditions of observation were controlled.

Wundt did not dirctly concern himself with the topic of this book. To Wundt precise introspection of the higher mental processes was impossible because (1) the experiment could not be repeated exactly, (2) introspection could not be directed to the phenomena to be observed, (3) attention could not be properly controlled, and (4) the conditions of introspection could not be systematically varied.

Many psychologists did not accept these strictures. Psychologists have always been solving problems that other psychologists told them were impossible. Binet, Külpe, Woodworth, and others went ahead, gave

their subjects thought problems, and asked for informal introspective reports. Some of the tasks used to stimulate thought were quite simple: "Give a word that means the opposite of compact." Some were complex: "What is the more delightful, the smell of a rose or its appearance?" "Can our thought apprehend the nature of thought?" Binet asked his ten-year-old daughter just to close her eyes and think of anything. Woodworth asked his subjects: "Should a man be allowed to marry his widow's sister?" [2] Syllogisms were also used.

THE WÜRZBURG SCHOOL

A concerted attack on thinking was begun about 1900 by one of Wundt's students, Oswald Külpe. Külpe and his followers at Würzburg asked their subjects to make judgments, e.g., which of two weights is heavier, and to report what went on in their minds while reaching the conclusion. Or, using the familiar association method, they presented their subjects a stimulus word, with instructions to respond with another word, then to give full introspective reports.

IMAGELESS THOUGHT. One outcome of this work was a controversy centering around the question of imageless thought. This controversy illustrates—as do so many other controversies in the history of psychology—the critical influence of the investigator's psychological and philosophical background. Observers trained by Wundt and Titchener, when they examined their thoughts, "found" sensations or images of sensations, even though they had to look hard to find them. The Würzburg subjects, like those of Binet in France and Woodworth in America, did not always find that their thoughts were made up of images, perhaps because they did not look hard to find them. Instead of palpable sensations and images, vague unanalyzable experiences were reported, which in Würzburg were called *Bewusstseinslagen* (usually translated "conscious attitudes"). One result of this controversy, discussed more thoroughly in Chapter 2, was a loss of faith in the introspective method. The controversy has lost its pungency, even though no one was officially declared winner, and a "final comment" on imageless thought was delivered by Woodworth (*1938,* page 788).

(*Einstellung.* A positive contribution of the Würzburg School, not completely understood at the time, has turned out to be the most important outcome. Previous work on reaction time had emphasized

[2] Titchener commented on this question in a footnote: "I may be obtuse; but I confess I can find in this question no food for thought" (Titchener, *1909*).

the preparation, or mental set, and the reaction itself was considered by several physiologists and psychologists as a "prepared reflex." Watt, an Englishman working with Külpe, applied this insight to the association experiment. Inventing a method that later became known as "systematic experimental introspection," he divided the association experiment into four periods: (1) the preparatory period during which instructions were given, e.g., "When the stimulus word is given, name an opposite," (2) the presentation of the stimulus word, (3) the search for the response word, and (4) the response. The subjects were asked to introspect on only one period at a time. The surprising result was that there was little to report in the third period. The work of thinking, as far as introspection could disclose, seemed to be done in the first or preparatory period. When the thinker is well prepared, the remainder of the sequence runs off automatically, with little conscious content.

Thus the emphasis shifted to the preparation. When the subject accepts the task, or *Aufgabe,* he is prepared to respond accordingly. Acceptance of the instructions establishes an *Einstellung,* or, as Ach called it, a determining tendency. The term mental set was often used as a synonym until the mental part of the term went out of fashion.

With another of his students, W. L. Bryan, Külpe showed how the *Einstellung* influences perception. They exposed nonsense syllables of different colors tachistoscopically, but varied the instructions. A subject might be asked to report what letters he saw, the number of letters, their colors, or the pattern formed by the syllables. The results showed that the mind could grasp that attribute it was prepared for, but missed the others. This was experimental psychology in a modern sense, more so than Watt's work. The independent variable, the instruction, was systematically varied and clearly related to the dependent variable, the report of perception. Neverthless, this type of experiment did not catch on at that time.

Later May used Watt's fractionation method with some improvements. During the preparation period his subjects reported kinesthetic and visual images, as of diagrams symbolizing the relation between the stimulus word and the response word, and also no images at all. During the third or main period they often had nothing to report, since the response came automatically, but sometimes they reported conflicting responses. The introspective data were objectively verified in one respect since response times for automatic responses averaged

about one second, while those for which conflicting responses were reported averaged about one and one-half seconds. (This introduces the principal use of the introspective method today. One gets a hypothesis or a subjective classification scheme from examination of his own thoughts, and then tries to devise an objective verification.)

THE SUBTRACTION METHOD. The reaction-time experiment suggests a method for analysis of mental processes which turned out to be unworkable. Donders, almost a hundred years ago, attempted to time higher mental processes simply by measuring the time of a complex reaction and subtracting the simple reaction time. Others went even further. The time required for cognition was inferred by measuring cognitive response time, as when many stimuli are to be differentiated, and subtracting the simple sensory reaction time. Judgment time was inferred by subtracting association time from total time, which included association and judgment.

The average times were not consistent, but the chief reason for the abandonment of this method was Külpe's argument against the elementistic assumption. It is true that complex reactions take more time than simple ones. But the change in instructions does not merely add another mental process; it changes the whole task. The mental performance is prepared in advance, and the stimulus word releases an integrated response.

SELZ AND THE SCHEMA. The Würzburgers rejected one assumption of associationism, the assumption that thoughts are always images, open to conscious observation. Selz, working with Külpe but influenced also by Kohler and the new Gestalt psychology, went further and rejected the associational mechanism. He presented the stimulus word and the *Aufgabe* at the same time on a card—e.g., Poem, Superordinate?—and asked his subjects for an introspective description of their mental activities. The results led him to stress, not the interaction of associations, but the integration of the *Aufgabe* and stimulus word into a *Gesamtaufgabe*. This consciousness of the total task guides the production of a complex of relationships, not just a class of words in a specified relationship to the stimulus word.

When the task is difficult, the complex that is integrated is not complete. The response word is missing, but it is schematized or anticipated in the total *Aufgabe*. Productive thinking consists essentially of completing this incomplete schema. Woodworth asserted in 1939 that

the anticipatory schema "lays claim to a position of fundamental importance in the psychology of thinking." [3]

THE DECLINE OF *Denkpsychologie.* This interest in *Denkpsychologie,* or the psychology of thought, during the first few years of this century had some of the characteristics of an exciting new movement that promises to grow and spread to other countries. The promise was not fulfilled. Külpe's interests changed. The First World War impeded scientific communication and Germany's prestige declined. Most important, the development of behaviorism cast suspicion on all introspective investigations, and this suspicion was generalized to a taboo on thought itself.

Denkpsychologie was not a blind alley; at least no one proved that it was. But other alleys to other goals appeared more attractive to active psychologists. The goal, understanding thinking, has always been an attractive one, but the revival of interest in this topic came from other directions and the research followed other approaches.

THE DEËMPHASIS OF CONSCIOUSNESS. Today the description of the contents of consciousness is carried on more by novelists than by scientific psychologists. Critical psychology's loss of interest in consciousness is, to the nonscientific observer, one of the most pathetic disillusionments of scientific progress. Astronomy destroyed man's belief that his world is the center of the universe; then anthropology showed him that his cherished customs and moral standards are not the only rules to live by. Now psychology tells him that his intimate thoughts, the contents of this personal consciousness that lies so close to his ego, have only a questionable scientific status and really have much less power to influence his behavior than their vividness and intimacy had led him to expect.

The usefulness of subjective reports of conscious experiences cannot be taken for granted but must be proved, like the usefulness of any other data. And this will be carefully considered in a later chapter.

GESTALT PSYCHOLOGY AND COGNITIVE PATTERNS

Gestalt psychology began in Germany with the study of perceptual organization. Wundt and the associationists had pictured complex perceptions as built up by the synthesis of elementary sensations. Wert-

[3] For a scholarly discussion and evaluation of the Würzburg School and of Selz, see Humphrey (*1951*).

heimer, a student of Külpe, objected, saying that the perceptual pattern is primary. One perceives a triangle directly, not by adding three sides and three angles. Perception is determined by the given sensory data and the dynamics of organizing forces in the brain. Because of these brain forces some patterns, called "good figures," have priority over others, just as a soap bubble is more likely to take the shape of a sphere than of a pyramid. When perception is modified, it is not usually by the accumulation of little changes, but by a reorganization of the perceptual field. When a new force acts on the soap bubble, it does not change a little piece of the sphere; rather, the whole field of forces is reorganized into a new equilibrium.

Gestalt psychologists do not differentiate thought from other activities. The laws of organization apply to all. Their early experiments dealt chiefly with perception and depended on subjective data, but Köhler applied the Gestalt approach in 1917 to the behavior of apes solving problems. His account of these experiments emphasized the perception of relationships, insight into the organization of the solution pattern rather than associative learning of new responses.

Wertheimer in 1920 (and again in 1945) described human thinking, emphasizing and recommending the reorganization of patterns rather than piecemeal attack on the problem situation. Good thinking consists of seeing the structure of the problem clearly, restructuring it as necessary. According to Gestalt psychology, those psychologists who treat thought as a subhead under learning, and thus try to study thinking by experiments on learning mazes and memorizing words, are doomed to failure. These experiments have artificially eliminated the possibility of good thinking by depriving the subjects of the opportunity to perceive the significant relationships.

This emphasis on organization of patterns raised questions about the structure of cognitive and social, as well as perceptual, fields. It was Koffka apparently, expanding the concept of figure and ground, who popularized the concept of frame of reference, which has been so much used recently in interpreting political attitudes and other judgments in a social context. Gottschaldt showed how patterns can be embedded in larger patterns, thus becoming harder to disentangle from the whole. Lewin's topological theory of personality has dealt seriously with regions of the ego and the life space, their differentiation and overlapping. Koffka, Werner, and others have interpreted the development of children's thought and action in terms of Gestalt

principles. Maier, as we shall see later, espoused a kind of Gestalt psychology in his studies of reasoning in rats and humans. Tolman has attempted to infer, by objective procedures, the cognitive maps of rats, thus restoring the outmoded (in America at least) term cognition to scientific respectability.[4]

PERCEPTUAL, CONCEPTUAL, OR COGNITIVE? Those aspects of Gestalt psychology that depend on subtle introspection, speculative brain dynamics, and the questionable notion of "good figure" have not been enthusiastically received in Great Britain or America. But all psychologists have been forced to recognize the importance of Gestalten, configurations or patterns, and to try to explain how they are organized and used. The importance of patterns was demonstrated also by many experiments on concept formation, at first by introspective procedures, later by objective procedures.

Probably everyone will admit that people organize the world somehow, differentiating in-group from out-group and cabbages from kings. Some who wish to stress Gestalt principles speak of this as perceptual organization. But it is not perception exactly, so others speak of cognition and cognitive organization. Others, mindful of concept formation, speak of conceptual organization. They all seem to be referring to the same phenomena, but the different traditions have not yet pooled their data or their terms.

THOUGHT AS AN INSTRUMENT OF ADJUSTMENT

The associationists began with the contents of the mind as introspectively observed. Darwin published his *Origin of Species* in 1859, and gradually, as the implications of evolutionary doctrine became clear, the emphasis shifted from the contents of the mind to the mind as an instrument of adaptation to the environment. Picturing man as an active animal, trying to hold his own in the struggle for existence, put the emphasis on psychological functions or processes rather than contents.

FUNCTIONALISM AND FUNCTIONS

As early as 1890, William James complained that the introspective analysis of thought into elementary contents was too static. In an oft-

[4] For a recent statement from Switzerland of the Gestalt point of view, see Meili (*1951*) on *Denkpsychologie*.

quoted passage he says: "The rush of thought is so headlong that it almost always brings us up at the conclusion before we can arrest it. . . . The attempt at introspective analysis in these cases is in fact like seizing a spinning top to catch its motion, or trying to turn up the gas quickly enough to see how the darkness looks."

John Dewey, like William James, was a psychologist who became a philosopher with a psychological point of view. His psychological and philosophical writings popularized the notion that man's mind and achievements, including rules of logic and systems of philosophy, were human solutions to intellectual and social problems. In an influential book called *How We Think* he analyzed reflective thought into five stages or processes that were widely accepted as a functional description of a "complete act of thought" (see page 27).

Famous thinkers have often been asked to describe how they made their discoveries or created their masterpieces. Helmholtz, the great physicist-physiologist-psychologist, had done this, as had Poincaré, the great mathematician. Graham Wallas, a British political scientist whose role in British intellectual life was similar to Dewey's in America, was interested in thought as a sideline. He wanted to help the average citizen control and improve his thinking, so he collected the contributions of the associationists and the Würzburg School as well as Dewey's analysis and the accounts of the great thinkers. In 1926, in a book called *The Art of Thought,* he systematically described the various steps in creative thinking as preparation, incubation, illumination, and verification (see page 27).

These descriptions of functions resulted from logical and subjective analysis. But experiments were soon undertaken to study and differentiate the higher mental processes. Ruger, for example, in 1910 described the activities of his subjects solving puzzles as analysis, conscious variation, and the setting up and testing of a hypothesis. In 1927 Hollingworth separated judging from perceiving on the one side and guessing on the other. Judging was done with more accuracy and confidence than guessing, but less than perceiving.

CONTROVERSIES IN COMPARATIVE PSYCHOLOGY

Controversy about evolution stimulated comparisons of the capacities of different species. Thus comparative psychology got its start, with some investigators trying to find similarities between men and the other animals, and others trying to find differences. Those who

claimed to observe higher intellectual activities in lower animals were often quite uncritical about it, so C. Lloyd Morgan laid down the rule that no performance should be attributed to a higher capacity, like reasoning, if it could be explained by a lower one, like experience or instinct. Morgan's canon was supposed to be a special case of the law of parsimony so often invoked in the other sciences.

Instead of relying on anecdotes and casual observation, Thorndike in 1897 put some cats in a problem box and watched them try to escape. He saw no evidence of ideas or reasoning. His interpretation favored exploratory activity and accidental success. Performance improved with practice because the successful activities were rewarded. This interpretation, which came to be known as *trial-and-error* learning, was stated in terms of associations, but the associations were between stimuli in the problem situation and overt responses, rather than between ideas.

In England soon afterward Hobhouse, claiming that Thorndike had not given his animals a fair chance, undertook a series of experiments with animals in a variety of problem situations, all of which were visible and rather simple. His observations led him to assert that the higher vertebrates could perceive relationships between objects and respond adaptively to them. Experiments of this kind were continued by Yerkes, Köhler, and others, stressing problem solving by *insight.*

Thus the argument between trial and error and insight was born. It has enjoyed a lively career for half a century, sometimes contributing methods and results to the progress of psychology in general, sometimes going off into special disputes of no interest except to the disputants. There is no need to follow the argument *per se,* but in one form or another these interpretations are bound to reappear in later chapters.

BEHAVIORISM AND ITS CONSEQUENCES

Working with uncommunicative animals, comparative psychologists and experimental biologists were forced to devise their own techniques. They were able to perform enlightening experiments on many psychological topics by strictly objective methods. John B. Watson, who began his psychological career working with animals, announced the behaviorist position in 1913 by declaring that psychologists should not use any other methods. For many years experimental psychologists,

including those not dedicated to any particular "ism," had demonstrated great ingenuity in inventing objective methods for attacking difficult psychological problems. The methods of Ebbinghaus in Germany and Pavlov in Russia, for example, had been quickly adopted in other countries, especially in the United States. Thus the call for objective study of behavior got a generally favorable reception.

The implications of the behaviorist approach to thought, however, did require some defense. When someone thinks, an outsider may not be able to observe any behavior at all, until the results of thought are announced. Watson got out of this predicament by equating thought to "inner speech." Speech is behavior, to be sure, acquired as an adaptation to social life. But social pressure and ordinary lassitude interfere with overt speech under many conditions. Thus the overt speech of uninhibited children migrates inward, as it were, to become covert or implicit in the socialized adult. Association between one idea and another in consciousness became, to the behaviorist, association between a covert symbol for a stimulus and a covert symbol for a response. Behaviorists began to define thinking as covert behavior, or behavior involving symbols.

To nonbelievers this seemed like obscurantism, pushing the problem inward beyond the possibility of empirical attack. But Watson and his followers believed that inner speech movements are not completely covert. Some efforts were made to record tongue movements during thought. Nothing very specific turned up because the recording instruments were not sensitive enough or because they were not attached where the thought was. The musculature does show more activity in general, however, when the subject is thinking of something than when his "mind is a blank."

We must conclude that the influence of behaviorism on the psychology of thought has been a negative one. Watson showed how the phenomena of thought could be rephrased to fit his system of behavior, but he did not show how to do anything about it. His criticism of subjective methodology scared many psychologists away from thinking. The distinction between thought as conscious content and thought as a problem-solving function (see page 19) was not always grasped, so the taboo against introspection spread to a taboo against the investigation of intellectual activities in general.

Modern behaviorists, more sophisticated than Watson in some respects, have avoided thinking, preferring to deal first with the simple

behavior for which their methods and principles are most suitable. The admission of hypothetical constructs to modern behaviorism, however, seems to open the door to the unobservables, such as anticipatory schemas and cognitive patterns, that the early behaviorists excluded.

THOUGHT ACQUIRES MOTIVES

One survival of the dualism of early psychology has frequently been expressed, and more frequently assumed, as a contrast between opposites: between instinct and intellect, or between emotion and reason. Comparisons between species, stimulated by evolutionary doctrine, at first reinforced this survival of faculty psychology: man has the power of reason; the lower animals are controlled by instinct or habit.

But evolutionary doctrine also saw man endowed with the same instincts as the other animals. And when objective problem-solving experiments were undertaken with human subjects, the necessity for motivating the subjects became obvious. The important relationship was explicitly stated by McDougall when he wrote that all of man's intelligence was at the service of the instincts, any instincts.

Once the artificial distinction between motivation and thought was overcome and the relationship between the two was apparent, it became necessary to work out the details of the relationship. Woodworth powered this movement in 1914, pointing out the connection between interest and ability. Freud, McDougall, and others speculated about the learning of new drives, or at least greatly modified secondary drives, on the basis of primary drives; and the modern behaviorists are attempting to show precisely how this learning occurs in some simple cases. Working in the other direction, the effects of various drives on attention, perception, imagination, thought, and judgment are an active topic of experimental psychology today, as we shall see in later chapters.

The Würzburg School discovered the need for dynamic concepts in understanding thinking and tried to fill the blank with the *Einstellung*. Because its only credentials were introspective and because psychologists were interested in other things, this concept did not catch on. Years later, after dynamic concepts have become more fashionable and objective methods for their investigation have been devised, the *Einstellung* by one name or another is a favorite topic of investigation. We cannot say that the Würzburgers were premature

dynamic psychologists (because they came *ex post facto* to the *Einstellung*), but we can use this story as an illustration, in the history of psychology, of what in the creative thinking of the individual is called incubation.

PATHOLOGY OF THOUGHT. The disordered mind of the madman has long been an intriguing subject of speculation, and many observers of the insane and the neurotic have attempted to contrast their irrationality with the thought of rational man. The gradual development of Freudian psychoanalysis in the first decades of this century helped to shift the emphasis from thought, in the sense of rationality, to motivation.

Because of Freud's associationist background, he used thought in its subjective meaning, as a conscious content. But his shrewd observations forced him to describe, or invent, mechanisms by which conscious thoughts are disguised and symptoms are formed under pressure of the censor, or superego. The mechanisms are thought processes in a sense: repression, fantasy, symbolism, rationalization. But the emphasis in psychoanalysis has always been on the drives behind thought, their detours and collisions, rather than on intellectual processes *per se.*

THOUGHT ACQUIRES A CULTURE

When the early anthropologists of our civilization, fascinated by the theory of evolution, began to study less literate peoples, they interpreted savage thinking as prelogical and irrational. A leader in this movement, E. B. Tylor, stated in 1871 that primitive man was animistic; that he looks at the world anthropomorphically and endows all things with a consciousness like his own. When anthropology gradually became more critical and empirical, the differences between the mind of primitive man and civilized man seemed to dwindle away. The final coup, perhaps, was delivered by Paul Radin when he collected many wise and witty sayings from preliterate cultures in a book by the flattering title *Primitive Man as Philosopher.* Anthropologists gradually came to state the differences in terms of cultural background, interests, and information rather than in the thought processes *per se,* though adequate comparisons between representative samples of different peoples have never actually been made.

The early social psychologists, writers like Tarde and Le Bon, were intrigued by man's irrationality in social situations. Illustrations of

imitation and mob violence were collected to show that social influences make people suggestible and that membership in a crowd reduces the individual's intellectual powers to the lowest level. When social psychology borrowed the objective methods of experimental and differential psychology, demonstrations of irrationality became very popular.

These bizarre overemphases were brought into better balance when the textbook writers, e.g., Gardner Murphy *et al.* in 1931 and 1937, undertook their undramatic function of integrating strange facts and diverse points of view. Social psychologists' primitive glee in showing that people are irrational, prejudiced, superstitious, suggestible, and undemocratic was muted by more systematic attempts to discover the origins and operation of values, attitudes, and opinions, whether sensible or foolish. Later chapters will show that research on bad judgment has contributed in a roundabout and incomplete way to the psychology of good judgment.

When experimental methods were applied to the activities of people in groups, intellectual activities were among the first to be studied. As early as 1920 Floyd Allport was comparing associations and judgments of people alone and in groups. About the same time H. T. Moore was showing the influence of majority and expert opinion on the judgments of his subjects. Others presented small groups with complicated problems and recorded, as well as they could, the social processes, communications, and recommendations that led to the solutions. The commercial development of convenient recording devices has accelerated this trend. Industrial psychologists, with a practical interest in improving group decisions, have recently been attracted to this type of investigation.

Just as important as the specific results of these developments in social psychology is the general effect of socializing all psychology. Coming out of his laboratory, the psychologist was occasionally asked by an anthropological colleague how his results would differ if the experiment had been done in New Guinea. Careful investigators began to consider the implications of their subjects' attitudes toward any experiment or test, the instructions given by the experimenter, the laboratory atmosphere, and the social significance of the responses. In the past these influences on thinking had been treated as nuisances, to be eliminated by careful control of the experimental con-

ditions. But a nuisance taken seriously often becomes the starting point of more adequate theorizing.

PSYCHOPHYSICS: FROM SENSATION TO JUDGMENT

Alongside the dramatic rise and fall of psychological schools, experimental psychology has been steadily expanding its interests and achievements. Introspective procedures have come and gone, but psychophysics since 1860, when Fechner published his *Elemente,* has been predominantly an objective science.

Fechner's interest at the outset was subjective, metaphysical, almost mystical. He wanted to measure the intensity of sensations so he could tie the inner world of consciousness to the outer world of physical stimuli with the firmness of a mathematical equation. As so often happens, Fechner's aims have been ignored, but his psychophysical methods are in daily use. Experiments were soon undertaken to determine the acuity of each of the senses, acuity being defined eventually as precision of judgment. Sensory capacities under different conditions were charted, and anatomical and physiological correlates were explored. Fechner's methods for studying preferences for physical stimuli, e.g., rectangles of various proportions, were expanded into a new science of experimental aesthetics.

When psychologists asked how these judgments were made, they naturally turned to introspection for evidence. Careful observers reported images and other conscious contents when comparing a sound or a weight with a standard. As in the *Denkpsychologie* experiments, the results were disappointing. When the observer was well prepared, the judgment seemed to come directly, as soon as the second stimulus was presented. Perhaps the most interesting results were those of Müller and his students about 1900. They noted that after a series of judgments the observer was not comparing the lifted weight with an image of the standard weight, but rather with some absolute impression of the whole series for which he had gradually become set or *eingestellt.*

Refinement of the psychophysical methods attempted to hold the judging process constant so that functional relations between some aspect of the stimulus and some perceptual consequence could be clearly determined. Under ideal conditions it was assumed that the judging process was standardized, and therefore psychophysical data

could be published as "from the left eye of Observer B." But in the effort to improve the psychophysical methods, a number of constant errors and other phenomena of judgment came to light. When these were taken as the objects of investigation rather than as incidental to the measurement of sensation, the experimental psychology of judgment was born.

An important constant error was often observed when each of a series of stimuli is compared with a standard stimulus. A midpoint or center can be calculated from the series of judgments, which under ideal conditions may be expected to coincide with the standard stimulus. It is called the point of subjective equality, or PSE. But it is not always exactly equal to the standard stimulus; the discrepancy is called a constant or systematic error. Hollingworth discovered about 1909 that when several series are used, the larger stimuli are underestimated and the smaller are overestimated. The PSE's for each series approach the center of all the stimuli judged. This phenomenon, which he called the "central tendency," and others called the "series effect," has been observed in many experiments with different kinds of stimuli.

It is not necessary to compare each stimulus with a standard stimulus. The rating method appeared about the end of the last century, applied first to the judgment of affective stimuli and other complex objects of judgment. When the rating method was used in a psychophysical experiment, e.g., on the loudness of sounds, it was called the method of single stimuli, since no standard was available for comparison, or the method of absolute judgment, since, as noted above, the judgments seemed to be direct.

Something similar to the series effect was noticed when the method of single stimuli was used, as in rating the affective tone of colors or odors. Judgment of each stimulus was influenced by the preceding stimulus, in fact by the whole range of stimuli. In the 1920's this phenomenon was called affective contrast and affective equilibrium. As Gestalt psychology extended its influence, the term frame of reference was loosely applied to these facts when they came to light in many fields of psychology.

During the First World War and afterward, applied psychologists attempted to derive some generalizations from these facts and others in the hope of improving the accuracy of judgments. The psychology of judgment has remained relatively insulated, however, from other psychological topics and, strangely enough, has been completely ig-

nored by the system builders. Later chapters will show that the ingredients are available for a systematic psychology of judgment, based on information from psychophysics and social psychology, which is not inconsistent with modern dynamic theories.

THE TESTING MOVEMENT

People have always speculated about differences between people, but differential psychology waited for progress in general psychology. The pioneers in this movement, Galton, Cattell, and Binet, investigated mental processes before they constructed mental tests. Galton and Cattell had been working on simple processes, while Binet had been working on reasoning, studying the thinking of his two daughters as well as chess players and arithmetical prodigies. Hence he characterized intelligence as ability to understand directions, to maintain a mental set, and to correct one's errors, and the tests he constructed along these lines were most useful in the school situation.

Once the testing movement got started, after the First World War, it acquired terminology, methodology, and, to some extent, personnel of its own. Hints for new tests soon came from successes and failures of previous tests as well as from the findings of general experimental psychology. In recent years the testing of intelligence and the psychology of thought have progressed side by side with only occasional interdisciplinary contacts.

The generalizations of differential psychology have for the most part been of a statistical nature, based on correlations between test scores. Psychological interpretations of the relationships have been secondary and tentative. It was noticed early that all tests that could be called intellectual are positively correlated. In 1904 Spearman announced his two-factor theory, stating that success in any task depends on *g*, a general ability required for all, and *s*, an ability specific to each. Later he devised methods for computing the amount of *g* involved in each test and the amount displayed by each subject. When he came to discuss the nature of *g*, he spoke of a fund of cerebral energy that is manifested in eduction of relations and of correlates.

The two-factor theory encountered many objections. In England group factors, intermediate between *g* and *s*, had to be recognized. In America Thurstone developed a more general method that analyzed the correlations into a number of factors, none of which need be more general than the others. This method results in many factors or

abilities, and a whole new methodology was developed for identifying abilities and purifying them in later tests. The nature of these abilities, like the nature of *g*, is not clearly disclosed by the analysis, but some have been assigned names of particular interest to the present discussion: fluency, reasoning, judgment.

The practical problems of education have encouraged attempts to test important intellectual abilities, such as critical thinking and recognizing assumptions. The demands of personnel selection likewise have promoted new kinds of tests, such as mechanical comprehension and research aptitude.

TESTS FROM THE CLINIC

Kurt Goldstein, a German neurologist who came to America, approached psychopathology from a Gestalt point of view. He found that the associationist concept of reflex, standard to most neurologists, was inadequate to explain or even to describe the symptoms of war veterans with head injuries. The troubles of these patients consist not of isolated traits but in coming to terms with the total environment. The intact healthy person can organize his perceived environment and reorganize it in accordance with his needs. The brain-injured patient is bound to one view of his situation and task. In other words the brain injury results in a fixed concrete attitude toward the world. The normal person can adopt one concrete attitude toward the materials at hand, organizing them one way, or he can adopt another attitude and organize them another way. The normal person has the capacity for abstract thinking, for shifting from one approach to another. With his students Goldstein devised tests of concrete and abstract thinking that have been widely used in psychological clinics.

Tests of concept formation or abstract thinking have been constructed by Ach (of the Würzburg School), Vigotsky, Weigl, and others. The Vigotsky block test, as revised by Hanfman and Kasanin, is used in many American clinics. As psychologists became dissatisfied with getting a single quantitative score from an intelligence test, they invented ways of scoring various subtests so as to get measures of abstract thinking.

The construction of a new, or presumably new, test is good clean fun, and tests are often constructed, named, and used without careful analysis. But the pertinent evidence that is available will be systematized in later chapters.

THE NATURE OF THE TOPIC

This brief review of the historical antecedents of our topic indicates that thought, like most other psychological concepts, has had different meanings at different times. Some clarification can be obtained, and the topic of the present book can be delimited, by examining three meanings of thought. They have often been confused, and it is important that they be kept distinct.

THOUGHT AS THE CONTENT OF CONSCIOUSNESS

This is the classical usage of the term in psychology. This is what people mean when they ask, "What's on your mind?" or "What are you thinking about?" or when they offer "a penny for your thoughts." These thoughts are observable only to the thinker of them; the method of investigation is self-observation. This method is adequate under optimal conditions, when simple demands are made of it. Under other conditions the method is quite unreliable and has in fact been rejected by many critical psychologists. The rejection of this meaning of thought has unfortunately, by failure of discrimination, spread to the following meanings also.

THOUGHT AS PROBLEM SOLVING

This is a very different meaning, referring to activities rather than contents. This is what people mean when they say, "I had to stop and think." The activities of problem solving are not usually observable, either subjectively or objectively, but may be inferred from behavior and reports of contents. Symbols, concepts, and cognitive patterns fit into this meaning of thought, not as the defining characteristic, but as materials that may or may not be used in problem solving. This meaning of the term, unlike the first meaning, is applicable to infrahuman animals. Thought as problem solving has been contrasted superficially with muscular action, but thought in this sense is clearly a verb: "I'll have to think about this problem." Approximate synonyms are reflect, ponder, deliberate, and cogitate.

THOUGHT AS JUDGMENT

When someone asks, "Do you think it is going to rain?" he is not asking for a report of your consciousness, nor is he asking you to solve a complicated problem. He merely wants you to communicate

the outcome or conclusion of your thinking: "What is your judgment?" The third meaning is a special case of the second because a judgment is a special kind of problem. The special feature of judgment, as the term is used today, is that it is communicated by a response that is one of a small number of specified categories of response: *yes* or *no, low, medium,* or *high, guilty* or *innocent.*

Much of the confusion surrounding the psychology of thought can be cleared up by differentiating these three meanings of thought and then discarding the first one. Psychology is no longer defined as the science of consciousness, and the study of what people think about is not central to the psychology of thought. Two recent books (Humphrey, *1951;* Vinacke, *1952*) have given considerable space to this matter; hence the present book will consider the thinker's reports of his thoughts only when they are helpful in understanding the other meanings of thinking.

Modern psychology is more interested in processes or functions than conscious contents. How people solve problems and how they make judgments are very live questions to which the bulk of this book will be devoted. All the important aspects of intellectual activity can be treated under these headings.

REFERENCES

Boring, E. G., 1950. *A history of experimental psychology,* New York, Appleton-Century-Crofts.

Dewey, J., 1910. *How we think,* Boston, D. C. Heath & Co.

Humphrey, G., 1951. *Thinking: an introduction to its experimental psychology,* New York, John Wiley & Sons.

Meili, R., 1951. Denkpsychologie, in *Handbuch der Psychologie,* ed. D. Katz, Basel, B. Schwabe & Co.

Murphy, G., 1949. *An historical introduction to modern psychology,* New York, Harcourt, Brace & Co.

Ogden, R. M., 1951. Oswald Külpe and the Würzburg School, *Am. J. Psychol., 64:*4–19.

Titchener, E. B., 1909. *Lectures on the experimental psychology of the thought-processes,* New York, The Macmillan Co.

Vinacke, W. E., 1952. *The psychology of thinking,* New York, McGraw-Hill Book Co.

Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt & Co.

Chapter 2

The Thought Processes: Methods for Analysis and Classification

> As far as we can ascertain, the intellectual processes themselves are the same for all normal human beings in all times and places.
>
> —RALPH LINTON

THE principal purpose of this book is an integration of present knowledge of the psychology of thought, aiming at breadth rather than the exploitation of any special theoretical approach. We must recognize at the same time that many of the advances in this field have been made by those who did have a special ax to grind, a favorite apparatus, method, or principle to advocate. Information from all sources, obtained by all methods, should be evaluated and integrated. The desired breadth of approach is fostered by the assumption of a naturalistic point of view, to which we may return for orientation and clarity when fogged by the minutiae of research data or the extravagances of controversy.

THE POINT OF VIEW. This naturalistic approach begins, of course, with the organism in its environment. Thinking, like all other activities, is performed by live animals, with multiple motivation and fluctuating efficiency. The intellectual output of even the most ponderous commencement orator flows from moist tissue that is continually nourished, oxygenated, and rinsed by body fluids throughout his life cycle. The problems set up in the psychological laboratory, as well as the problems encountered by the organism in its daily tour of

duty, are only episodes in a complex career of adjustment to changing circumstances.

The problems that the lower organisms run into during life on the surface of this earth are chiefly spatial problems, locating places and objects within the environment, manipulation of objects, and locomotion from one place to another in spite of barriers. In these instances a large part of the problem-solving activities are overt, visible to the observer, and thus rather easily recorded. Such records of activity in the environment constitute the data for a large class of investigations of problem solving.

The environment that human beings operate in is more highly diversified. Human beings attach special importance to certain places and objects, as the other animals do, but, to a far greater extent than any other species, they do business with people and with symbols. A large fraction of anyone's day is spent in interaction with other people in direct and subtle ways, and from this interaction a large fraction of one's problems arises. Those aspects of the environment with symbolic significance, especially words coming from other people, either directly or via the printed page, also take on extraordinary importance, and many problems are of this abstract or symbolic variety.

Furthermore, these abstract representations of pertinent aspects of the environment, commonly called ideas, are not tied down to specific locations. They are detachable and portable. If a man has the problem of locating an outdoor fireplace in his small back yard, he may struggle with it "in his head" while riding home on the bus. He may even make rough drawings on the margin of his evening paper. Arriving home, he may find that his imaginary or mental solution is actually helpful when applied to his physical environment. In such cases it is a reasonable hypothesis that the problem-solving activities are carried on in an imaginary or ideational environment. Portions of the physical world are abstracted from the whole, remembered, and utilized at the thinker's convenience.

The consequence of all this is that the psychology of thought and judgment must be concerned with the motivated organism operating in and with a highly variegated environment consisting of people as well as things, of ideas as well as concrete objects. The organism's responses will be words as well as deeds and—methodologically most important—a large part of his problem-solving activities will not be visible to the observer. Studying these invisible activities is a central

difficulty in the experimental investigation of thinking. Neglecting them, or any other likely possibilities, would be absurdly unscientific.

THE NECESSITY OF ANALYSIS. Thinking is a fluid enterprise. Like raindrops coursing down a windowpane, intellectual operations coalesce and separate and run together again. A first approach to our topic could be a superficial correlation of occasions for thought and outcomes, ignoring intervening operations. Methods appropriate to this approach have been borrowed from other areas of psychology and need no special discussion here; results obtained by this superficial approach will be summarized in the proper places. The special character and the difficulties of the methodology of thought and judgment arise from the attempt to get below overt behavior and investigate invisible operations or processes assumed to produce the outcome.

Just as biologists find it convenient to analyze an organism's adjustment to its environment into such functions as respiration and digestion, psychologists often speak of intellectual functions such as imagination and decision. Early psychology accepted introspection as direct evidence for identification of mental processes. More sophisticated recent theorizing finds its rationale in such notions as intervening variable and hypothetical construct. The terminology is not important, nor is the philosophical argument which supports the theorizing. What is important is the procedure by which thought processes are identified, analyzed, and classified, and the assumptions on which the analysis rests. Analysis is a first step in any scientific endeavor, and an unproductive path may beckon as invitingly as a productive one. The methods by which hypothetical thought processes are inferred from observable data, then analyzed and classified, must, therefore, be critically examined and evaluated.

The difficulty of differentiating one intellectual function from another, describing and naming each, has often been emphasized, and it is partly this initial difficulty that has steered many psychologists away from this topic. But the difficulty is not peculiar to the higher mental processes; it is not easy to separate sensory processes from perceptual processes, nor motivation from emotion. Since the methods of analysis and the variability of the results are similar in the present area of psychology and in psychology as a whole, it is not too much to hope that a careful study of the possible ways of sorting out the thought processes will contribute to the larger endeavor.

THE DESIRABILITY OF BREADTH. Many attempts have been made from

time to time to differentiate one psychological process from another, e.g., insight from trial and error, but the proper goal, however remote, is a system of concepts that includes all intellectual activities, defines each, and clarifies their relationships. Within such a system petty antinomies may be more judiciously scrutinized.

One way of approaching the desired comprehensiveness would be to follow a few thoughtful, communicative people all day for a few days, attempting to note and classify all the intellectual operations that they perform. The labor would be tremendous and problems of observing and recording would certainly arise, but the effort would be broadening.

Another approach, more immediately fruitful, has been employed as the initial approach to the analysis of personality by Allport and Odbert (*1936*) and by Cattell (*1943*). Leaning heavily on the wisdom of tradition, they begin with the words used to characterize personalities. Thus the dictionary becomes the starting place. More critical methods carry on from there. In support of this procedure is the fact that there is nothing recondite about personality—or thought. The thoughts we think today as we read the daily paper may be as modern as jet propulsion, but the thinking of them is not. There are no undiscovered specimens awaiting the development of new instruments or techniques. The basic data have been observed and discussed by sages, poets, playwrights, politicians, and lawyers for centuries. Probably little of any consequence has escaped the attention of the wordsmiths or stumped their inventiveness. No comprehensive search of the dictionary for terms referring to intellectual processes has been made, but it is relatively easy to assemble a list of terms commonly used in ordinary language and in the professional literature of our topic. Such a list of nouns and verbs will indicate the breadth of the topic and some of the problems of classification:

abstraction	concept formation	generalize
analysis	conclude	guess
association	controlled association	hypothesize
attention	daydreaming	ideation
autism	decision	illumination
believe	deduction	imagery
cognition	deliberation	imagination
comprehension	fancy	incubation
concentration	free association	induction

inference	perception of relations	resonance
insight	planning	restructuring
interference	preparation	revision
intuition	problem solving	search
invention	productive thought	set
judgment	reasoning	thought
knowledge	recall	trial and error
locus analysis	reflection	understanding
orientation	reorganization	verification

Analysis of the intellectual processes deals with concepts like those listed above and ends with a separation of one process from another, e.g., recall from inference, insight from trial and error, or ideally with a number of concepts systematically arranged. In general there seem to be three methods by which psychological processes, including thought process, are sorted out, identified, and classified: subjective procedures, objective experimentation, and analysis of individual differences.

SUBJECTIVE PROCEDURES

Several subjective procedures have been used. They cannot be sharply distinguished from one another, but it is profitable to discuss them separately, pointing up the minor differences and evaluating the usefulness of each.

INTROSPECTIVE ANALYSIS OF THE CONTENTS OF CONSCIOUSNESS

As stated earlier, thought has long had a content definition, in terms of what a person thinks *of* or *about*. If this definition is accepted, the method of choice is systematic introspection, careful observation of the contents of one's own consciousness. In the history of psychology, introspective is not synonymous with subjective. The introspective technique is a special kind of self-examination which requires knowing what to observe and what to ignore. Practice and concentration are necessary.

This method has not been productive, however, except in the negative way mentioned in the previous chapter. The emphasis on contents is misplaced. A more important limitation is the lack of agreement that shows up when different observers are compared. The reason for this embarrassing lack of interobserver agreement is clearer now than it was 30 years ago. Research has demonstrated how, when a situation is ambiguously perceived, the perceiver's set, based on previous training, present motivation, and the like, plays a prominent part in deter-

mining the results. Since standardization of observers and their sets in this respect is neither possible nor desirable, this method in its classical form has little to offer modern psychology.[1]

LOGICAL ANALYSIS

One kind of analysis that is common in psychology and in many other branches of knowledge may be described as a logical or rational procedure. Logically, for example, it is clear that deduction is different from induction. Obviously, generalization and induction are similar and can be conceived as instances of the same psychological process. And so on. In the case of the thought processes this logical analysis is usually aided by self-observation. The thinker examines his intellectual activities to see if his logical distinctions hold and if his terms are apt. If he is scholarly as well as thoughtful, he will study the writings of other thinkers for possible verification or improvement of his classification scheme. If he can handle his own thoughts and the literature on the subject by means of a few related terms, he is likely to retain these terms and use them in discussing the thought processes.

CRITERIA OF CLASSIFICATION. Inspection of the terms listed above and published classifications of such terms indicates that the thought processes can be logically named and classified according to several overlapping criteria.

Some processes are classified in respect of the logical significance of the activity: induction, deduction, eduction, reasoning, inference, verification, hypothesis formation.

The basis of classification may be some characteristic of the product of the thinking. The term creative thought, for instance, seems to imply that the outcome is new, either to the thinker or to society. Other examples are decision, generalization, planning.

Some terms describe a process by indicating its relation to a larger activity: conclusion, preparation, revision, locus analysis.

Other terms have a principally negative meaning, indicating that some result has been achieved, the steps to which are not apparent: incubation, intuition.

Certain terms, not listed above, are adjectives which designate a special characteristic of thinking: autistic, controlled, disconnected, flexible, imaginative, rigid.

[1] For a concise history of attempts to define cognitive processes in terms of consciousness, see Leeper (*1951*).

SUBJECTIVE ANALYSIS OF THOUGHT PROCESSES

It is possible to examine one's own thoughts psychologically rather than logically, with the emphasis on processes rather than contents. This procedure will be favored by those who give thought a problem-solving meaning rather than a content meaning, trying to trace what the thinker *does* rather than what he thinks about. Naturally this approach is more congenial to functional psychology than analysis of conscious contents. John Dewey's (*1910*) analysis may be taken as a familiar example. Dewey's purpose was to analyze reflective thinking for schoolteachers, preparatory to explaining how thought may be educated. It is a subjective analysis, based, he says, upon "descriptions of a number of extremely simple, but genuine, cases of reflective experience," which were turned in by students.

> Upon examination, each instance reveals, more or less clearly, five logically distinct steps: (i) a felt difficulty; (ii) its location and definition; (iii) suggestion of possible solution; (iv) development by reasoning of the bearings of the suggestion; (v) further observation and experiment leading to its acceptance or rejection; that is, the conclusion of belief or disbelief.

Dewey's analysis is logical, to be sure, but it was no doubt intended to correspond to actual psychological processes. The fact that it has been often quoted by teachers of education and psychology argues that it must be convincing, subjectively at least, when teachers apply it to their own thinking. There seems to be no other evidence for or against it.

Graham Wallas (*1926*) analyzed the thought processes in much the same way. His purpose was to aid education and control of thought, and his method was a subjective analysis of his own thoughts, reports of his students and friends, and published anecdotes. Some of the processes he described had been described earlier, e.g., by Helmholtz. Wallas' contribution was to line these phases of thought alongside each other, describe them more clearly, and give each a distinctive name.

The preliminary assembly of information he called preparation. The second phase, called incubation, is described negatively. After the thinker has loaded himself up with the problem he lays it away for a while to hatch, and, though he does not consciously notice anything, something must be going on because later at some casual moment new, happy thoughts occur. The third phase, the occurrence of these happy

thoughts or new promising leads, he named illumination. The final testing of the leads is verification.

This analysis, like Dewey's, is superficially convincing. Dewey states that he is analyzing reflective thought, "the complete act of thought," while Wallas is particularly concerned with creative thought; but the difference is only nominal, and both illustrate their analyses with the same kind of examples. These two formulations of the thought processes may be taken as thoughtful, scholarly, but primitive attempts at subjective analysis.

A little evidence in support of Wallas' formulation can be brought forward. Lowes (*1930*) attempted to reconstruct the thought processes that produced *The Ancient Mariner* by studying Coleridge's notebooks and sources, and decided that his results were consistent with Wallas' descriptions. Such evidence is very thin, to be sure, but it indicates that Wallas' categories are at least not obviously unusable by other analysts. (The supporting evidence of Patrick's studies will be described below.)

The ubiquitous questionnaire offers another way of gathering subjective data on the thought processes. Rossman (*1931*) and Platt and Baker (*1931*) used questionnaires to check on the incubation phenomenon. They got positive answers from most respondents, an outcome that does not supply conclusive proof but does show that the respondents could understand what they were asked and could use the categories suggested to them. Perhaps with further refinement the questionnaire could be made to yield more convincing results.

THINKING ALOUD. These subjective analyses of thinking thus far discussed are one step from merely logical analysis toward actual psychological activities. The next step is to get records of the progress of thought. Perhaps it cannot be done, but the thinking-aloud or *réflexion parlée* method comes close to it.

As Claparède (*1934*) put it, introspection calls for a description of elementary mental contents, of sensations, images, and feelings, while the thinking-aloud method merely asks the subject to talk while he is working, reporting what he is doing, what he is looking for, what he plans to do next, and so on. It demands less of the subject than introspection, and the subject need not be highly trained. Thus it is more superficial, in a sense, but also more natural. The subjects usually talk about many things, including irrelevant comments and asides, expressions of joy and perplexity, as may be seen in the protocol on page 32.

The thinking-aloud method has been used often to identify one

process of particular interest or to compare the importance of two or more processes, e.g., insight and trial and error. Catherine Patrick's (*1935, 1937*) studies of creative thought go farther and attempt to classify the thought processes, as disclosed by the thinking-aloud method, according to the four-part analysis of Wallas.

Her method was quite simple and straightforward. In the first study she had 113 subjects, 55 poets of recognized ability and 58 nonpoets, write poems under her supervision and observation. A picture of a landscape was used as a standard starting place, and instructions were given to "write a poem about this picture or anything that you happen to feel like as you look at it."

Patrick's treatment of her shorthand records illustrates one way of using Wallas' categories and at the same time gives a more clear-cut meaning to these categories. As a sign of preparation she used the "thought change," that is, "any modification of thought, which was sufficient to cause it to form a new sentence, whether it had actually done so or not in the report." Over half of these thought changes occurred during the first quarter of the creative effort. During this first quarter, presumably, the poet's ideas are not yet dominated by any coherent theme; he is assembling his raw material.

The first formulation of a line of poetry was taken as the mark of illumination. This phase occurred most commonly in the second and third quarters of the creative effort.

Wallas' final phase of verification becomes, as Patrick extends it to the case of poetry, revision or elaboration. As such it was easy to identify by the rereading of the poem, the changing of a word or two, etc. Sixty percent of such revisions occurred in the last quarter of the experiment.

The study of creative thought in artists was conducted in the same way. Artists and nonartists, 50 of each, were given a selection from Milton's *L'Allegro* and asked to draw a picture about it or whatever it suggested. They were encouraged to talk about their efforts while the experimenter recorded both their words and their work. The criteria of the stages of thought were the same in general as for poetry. The first drawing of the rough outline of an object was taken as the mark of illumination. Shading an object to clarify it, addition of detail, or critical inspection of the picture as a whole was considered as revision. The results were practically the same as for the poets. The thought changes, indicating preparation, come first; the initial for-

mulation of an idea, taken as the mark of illumination, comes next; and revision comes last. Patrick concludes that her results support Wallas' description of these three stages of creative thought. (See Figure 2.1.)

The evidence for incubation is less direct since it must be inferred from the records, and just when it occurs cannot easily be determined. According to Patrick, "the presence of incubation was shown if an idea occurred early in the report, recurred one or more times, the subject meanwhile talking of other things, and at last appeared as the chief topic of the poem." By this definition incubation could be inferred

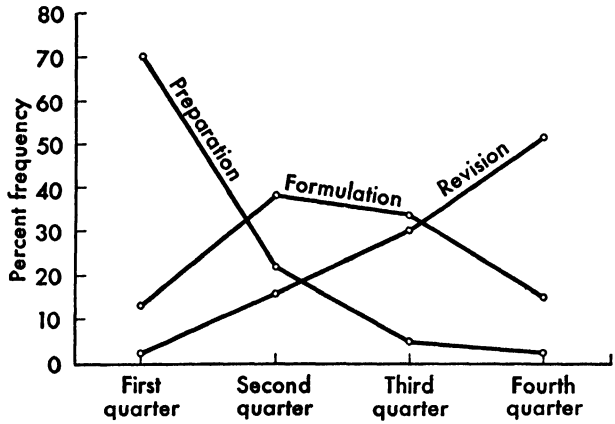


FIGURE 2.1. Intellectual activities of poets and artists during four quarters of a creative enterprise. Data from Patrick (*1935, 1937*).

with some certainty in over three-fourths of the protocols from both studies. Furthermore, when the experts were asked about their usual methods of work, about three-fourths stated that they incubate their ideas for a while before beginning.

These two studies by Patrick illustrate the collection of data by the thinking-aloud method—supplemented somewhat by observations of behavior—the subjective nature of the analysis, and the kind of results achieved. The protocols have usually been taken down in writing, often by shorthand, but many laboratories now have electrical recording and reproducing equipment well adapted to this procedure. Similar procedures have been used by Bloom and Broder (*1950*) and Burack (*1950*).

To facilitate the thinker's communication of his intellectual activities, it is necessary to put him at his ease and encourage him to talk freely. Many subjects like to think silently, voicing only the outcome of their efforts. An example and some preliminary practice are helpful if not prejudicial. Preliminary practice on word problems that are easily verbalized may get the subject off to a loquacious start. Patrick gave her poets a picture as a starter, and her artists a poem. Presumably, the thinking-aloud procedure is easier if the problem is one of words or numbers than if it deals with spatial relations.

Observation of Behavior. Thus far we have been discussing the experimenter's subjective analysis of the subject's subjective reports. It is possible also to record the subject's overt behavior and to apply the same kind of subjective analysis to these records, identifying processes, differentiating one process from another, aiming perhaps at a systematic classification of everything the subject does. This is especially feasible if the problem is a spatial one, such as ring-and-string puzzles, finger mazes, and geometrical problems, which stimulate the subject to make many observable movements. The movements may be described in words. Drawings, snapshots, and movies have also been used. Since verbalization is not essential, this method has been used with non-literate animals and children as well as human adults.

The classical descriptions of animal problem solving by Thorndike and Köhler will serve as familiar illustrations of this method. These experimenters, and others who have followed them, have aimed, not at a complete classification of problem solving, but at a characterization of the problem solving in respect to a controversial issue, usually the issue of insight versus trial and error.

When the problem solving of human adults has been under investigation, behavior observation and thinking aloud have often been combined. Durkin's (*1937*) study may be taken as a good illustration of this combination of methods.

Durkin had college students assemble pieces of board to make geometrical patterns. Each was instructed to formulate a plan and tell it to the experimenter before moving any pieces. When the subject finished, he was asked to retrace his solution verbally. Snapshots were taken at critical moments. A sample record follows. (M refers to manipulation. Dashes indicate pauses. E and O refer to experimenter and subject respectively. E * means that E asked O what he was thinking about.)

SAMPLE PROTOCOL FROM DURKIN (*1937*)

Plan: 2 minutes

1. Oh, my word (laughs in an embarrassed way).
2. — — — — I don't see anything — — —.
3. — — — can make no plan till I touch them — —.
4. — — I'll have to move them.

Solution: 8 minutes

26 Moves	*Remarks*
Remarks 1–4 are accompanied by M. during the long pauses.	1. — — — — E * — — "I can't talk — — —."
	2. — — — don't see how — — — —.
At 5 lifts one up into the air and examines it (M.).	3. — — feel warm — — — laughs — — —.
Another series of quick moves, one of which is right but not recognized (M.).	4. — — — have no idea — —.
	5. — — must stop and try to think.
	6. — — — don't think I can do it.
	7. — — — — feel ashamed — — —.
Another series of rapid moves (1 correct unrecognized.) (M.).	8. — — E "you had it once" O "Yes, there were 2 that went together perfectly — — but can't repeat it."
Puts 2 together wrong but adds third correctly, recognizes it, and very slowly; the last 2 moves are correct (foresight).	9. — — — — — E * "Just sitting here stupidly just moving pieces — — a vain attempt to do it without thinking — — —."
	10. — — — thinking mostly that I'm stupid.
	11. — — — — Oh, I see it now (recognition) (after the correct move occurs for the 3rd time).

Retrace:

I didn't think — — just luck — — found two edges of the same length on a slant. Now I can see that if you stopped for a plan, you could see that they were cut the same way; if you arranged the straight edges, the points would fit together inside. Did last slowly because thought it didn't fit accurately enough. (The retrace gives a hint as to how deep the penetration at the moment of recognition was.)

Durkin was able to classify many of her records as instances of trial and error or of sudden reorganization, but found it desirable to add another category, gradual analysis. Although transitional forms were

recognized, these three kinds of problem solving are differentiated in Table 2.1.

TABLE 2.1. Durkin's Differentiation of Three Thought Processes [a]

Trial and Error	Sudden Reorganization	Gradual Analysis
1. "Blind" groping	Groping suddenly stopped	No groping but a gradually developing understanding
2. Hindsight	Sudden foresight	Foresight
3. Confusion till the last moment	Confusion suddenly cleared	Cleared step by step
4. Hopeless feeling	Excitement, elation, sometimes relief	Satisfaction
5. Aim—to match pieces	To look for wholes or interrelations	To satisfy goal needs
6. Attention to goal distant, diffuse	Attention not centered on goal	Attention concentrated on specific goal needs
7. Attitude not definite, but wandering, haphazard	Passive, receptive	Active, directed search
8. Error curve irregular, may not drop out after solution; transfer poor	Curve irregular, then sudden drop; transfer good	Error curve step-like; transfer good
9. Manner baffled	Baffled, then suddenly well organized, efficient	Calm, well-organized

[a] From Durkin (1937).

EVALUATION OF SUBJECTIVE PROCEDURES

The value of these various subjective procedures depends on their purpose.[2] If the purpose is merely to obtain a convenient arrangement of topic headings under which the literature can be reviewed and educational or research recommendations can be made, the test is one of adequacy of communication of the available information. Analysis of this sort is often responsible for the plan of a textbook or systematic treatise, and the adequacy of the procedure for this purpose might be called its "literary validity."

If the subjective classification of the thought processes is for the purpose of obtaining a system of categories useful for theory and experi-

[2] For additional classifications of the thought processes from various points of view and for various purposes, see Blumenfeld (*1931*), Symonds (*1936*), Reitz (*1950*).

mentation, the first requirement is that the thought processes be clearly identified. A system of categories must be devised such that the data—in this case the thinkers' subjective reports of their intellectual operations or the experimenter's records of their behavior—can be sorted into one category or another with little overlap and little remainder. To classify comments or behavior as preparation or insight or revision is obviously a subjective procedure and, regardless of the acumen of the investigator, the reliability of the procedure must be checked. Some of the controversy over insight versus trial and error is no doubt due to neglect of this elementary principle. What one observer calls by one name another observer may call by another. Throughout the long history of the analysis of the thought processes, though many investigators have commented on the difficulties of analysis, none has presented any data on the usability of his classification scheme. And none has taken the simple step of enlisting a second investigator to classify the material so as to get a measure of reproducibility or reliability of classification.

Looking backward from our present vantage point, it now appears that these early attempts at analysis of thinking were handicapped by the assumption that the thought processes would be definitively revealed in subjective or objective events as observed by an astute psychologist. The role of subsequent interpretation and analysis in the very definition of the object of study, and the pragmatic nature of the ensuing categories, have only recently been understood as modern statistical logic has been applied to some of these ancient and subtle problems.

A Modern Technique for Analysis of Protocols. Obviously, the first test to be met by any up-to-date method of subjective analysis of protocols from a thought experiment is the test of interanalyst agreement or reliability of classification. The categories, whether called preparation, spectator behavior, hypothesis testing, trial and error, or whatnot, must be explicitly defined in respect to the material available for analysis; and the classifiers must be trained to use these categories, so that two or more classifiers, working independently, will classify the material in approximately the same way. The most defensible definition of the problem-solving processes then becomes the operations by which the material is collected and analyzed.

This requirement is nothing more than the common requirement for classification in other areas of psychology and social science, often

under the name of content analysis or coding. A set of categories is developed for the analysis of freely expressed opinions, personal documents, or responses to projective tests; then the analytical usefulness of these categories is tested by the agreement of independent classifiers. A statement of the reliability of identification of one process is meaningless except in relation to the alternative categories.

Somewhat closer to the analysis of the thought processes is a functional classification by McClelland and Apicella (*1945*) of the verbal responses of 14 college students to a frustrating situation. Verbal output was recorded, 131 responses in all, and later classified under four main headings: withdrawal, attack, limitation, and substitution, with a few subheadings under each. After discussing the classification scheme thoroughly, the two experimenters were able to classify the subjects' remarks with high reliability.

A study of group problem solving by Heyns (*1948*), using methods worked out by Guetzkow (*1950*), reached the conclusion that different analysts can break up a record into units with high agreement and that the functional classification of each contribution is moderately reliable. The categories used were such as goal setting, problem proposal, information seeking, and information giving. (For other examples, see Chapter 12, "Intellectual Activities of Social Groups.")

After it has been demonstrated that a set of categories can be applied to a protocol by two or more investigators, the question of breadth of application can be considered. Some utility will be granted to a system of concepts which can be applied by two analysts to a record of the thinker's behavior or of his comments. If the same set of concepts can be used for both records, its value is broader. Even more can be claimed for it if the system applies to records of the solution of a range of problems: verbal, numerical, and spatial, creative and routine.

Demonstration of interanalyst agreement in the use of a system of concepts is, of course, a first step, which may be called a demonstration of reliability. The next step, analogous to validity, is to test whether the analysis can be profitably related to data obtained by other procedures. If a concept such as preparation will stand up under a check for reliability, can the preparation period be timed? Can we determine the conditions which lengthen and shorten the preparation period? How is this period affected by the instructions to the thinker, by the thinker's interest in the problem, by his fear of failure? Once such thought processes can be reliably identified and marked off from other processes,

the way is clear for systematic experimental variation of the conditions of thinking, tying the results of these subjective procedures to data of a more objective sort.

OBJECTIVE EXPERIMENTATION

Superficially, there seem to be many ways by which experimental psychology separates one process from another, but they all reduce to one. Two activities may be considered functionally different if they follow different laws. The differentiation may be qualitative or quantitative.

QUALITATIVE ANALYSIS. In the qualitative experiment, evidence that two different processes are taking place is the occurrence of different kinds of behavior. If two subjects exhibit observably different behavior —e.g., one manipulates objects, the other sits and looks, during the same experiment—or if one subject exhibits observably different behavior at different times, different psychological processes may be inferred. The behavior may be observed and recorded at any level of complexity, e.g., turning right as contrasted with turning left, or multiplying as contrasted with dividing. But the different kinds of behavior must be unambiguously observably different. Otherwise the differentiation of the two kinds of behavior is a subjective procedure for which a reliability check is required. In addition to differentiating two or more psychological processes, it is desirable to determine the antecedent conditions that set each in motion, thus tying each process down at both ends.

QUANTITATIVE ANALYSIS. In the quantitative experiment some continuous dependent variable, such as time, frequency, or magnitude of some clearly defined response, is recorded, and the differentiation becomes a mathematical one, apparent when the values of the dependent variable are plotted on the independent variable or when an equation relating the two is written. Since this quantitative analysis is not commonly used in psychology, two illustrations will be presented, the first of which is a theoretical discussion, the second both theoretical and empirical.

Attempts to differentiate trial and error from insight by crude observation of kinds of behavior were not convincing—for reasons discussed above. Therefore several investigators turned to rate of progress, recording time and errors, as offering a more clear-cut separation. If, for example, an interpretation in terms of a trial-and-error process led

to an increase in time of solution while an interpretation in terms of insight led to a decrease in time of solution, it would be easy to design a crucial experiment. Unfortunately, both interpretations predict a decrease in time and errors. Hence the objective differentiation must be more subtle.

The shape of the curve of improvement was then offered as the distinction between trial and error and insight. A mathematical argument on this basis has been thoroughly worked out by Thurstone (*1930*). He considers the mastery of a task composed of a number of acts, each of which can be performed successfully or unsuccessfully, assuming that the probability of success on the task will increase with practice as errors in the component acts are eliminated. Plausible assumptions permit him to develop a theoretical learning curve of rather general applicability, which includes a constant, k, for the effect of making an error on the probability that it will be eliminated. That is, when k is small, performance of an erroneous act decreases only slightly the probability that it will be repeated; but when k is unity, an error once made will be eliminated completely. From the mathematical discussion it comes out that k may be larger than unity, in which case elimination of one error eliminates other errors not yet made. This may be considered rational learning or insight, turning the curve of improvement sharply upward toward errorless performance. If tasks fulfilling Thurstone's conditions could be set up, trial and error could be objectively differentiated from insight, or at least a variety of performances could be scaled in respect to k.

Actually the objective differentiation of trial and error from insight in terms of the shape of the curve has not materialized in the past two decades, probably because of the difficulty of setting up the tasks in measurable form. Tasks which require the accumulation of simple elements learned one at a time are easily arranged, and in fact it is such tasks that are most often used in quantitative research on learning. But it seems to be difficult to set up tasks that are mastered by insightful learning, as defined by Thurstone's k, and at the same time yield a learning curve of sufficient length. Such tasks are usually not mastered at all or are mastered very quickly.

The dilemma is not inescapable, however. Reed (*1946*) has devised tasks in which the learning of a concept automatically facilitates correct response to several groups of words; hence, when number of words learned is plotted on practice periods, the learning curves are

either linear or sharply accelerated. And Harlow (*1949*) has devised tasks which his monkeys master slowly at first, then, with continued practice, almost at once. Mastery of such tasks as these could perhaps be treated mathematically so that a parameter describing rate of mastery could be taken as characterizing the process underlying the mastery.

Recent experiments on the discrimination of numerousness furnish an illuminating example of the experimental separation of two processes previously assumed to be one. Taves (*1941*) exposed various numbers of dots, from 2 to 180, to college students for 0.2 seconds. They were instructed to estimate the number seen, and in general the

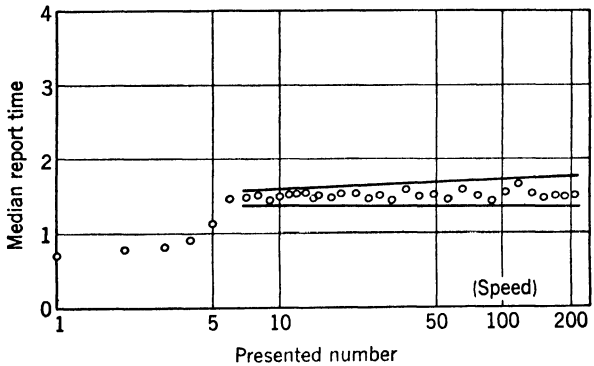


FIGURE 2.2. Differentiation of psychological functions. Time required for reporting the number of dots briefly exposed on a screen follows one law up to 6 dots and another law beyond 6 dots. Five subjects; 21 reports each. From Kaufman, *et al.* (*1949*).

number reported varied more or less with the number exposed. But scrutiny of the graphs of number reported on number exposed disclosed some interesting results. When 2, 4, or 6 dots were exposed, 2, 4, or 6 dots were reported. Variability was practically 0. Confidence was maximal. From 8 dots up, however, variability increased and confidence dropped sharply. Results from the bisection method also suggested that the numerousness of stimulus fields of more than 7 dots and of less than 7 was estimated by different operations. Taves concluded that "there are two mechanisms for the discrimination of visual numerousness."

Kaufman and others (*1949*) followed up this finding and confirmed it in more detail, instructing four subjects for accuracy and five for speed. The results for report time under speed instructions are shown in Figure 2.2. Other plots, showing report time under accuracy instructions and confidence under both instructions, are similar.

To prove that the discontinuity in the curves is genuine, Kaufman and his colleagues tried several logarithmic and hyperbolic transformations of the data. No treatment was able to make either branch of the graph an extrapolation of the other. A discontinuity was always apparent in the neighborhood of 6 or 7. When the same task was given with no time limit, the subjects subitized up to 5 or 6 dots and counted the rest (Jensen *et al., 1950*). It seems quite clear that two mechanisms or processes or functions are involved in this task, one operating above the point of discontinuity, the other operating below this point.[3] The authors recommend that the term *estimating* be applied only when the number is greater than 6, and propose a new term, *subitizing,* when the number is less than 6.

COMBINATION OF SUBJECTIVE ANALYSIS AND OBJECTIVE EXPERIMENTATION. It is possible to combine the subtlety of subjective procedures with the objectivity of controlled experimentation. In this respect the experiments of J. R. Patrick and N. R. F. Maier are instructive.

Patrick (*1934*) used a problem box for college students similar to one previously used by Hamilton for several species of animals. The subject had to choose one of four doors by which to escape from a small room. Following Hamilton, the experimenter classified problem-solving activities, according to sequence of responses, into the five types of Figure 2.3.

This is, of course, a subjective classification of activities and no evidence of reliability of classification is given. A reliability check by an additional classifier would probably show that the classification is feasible, but might lead to sharper differentiation between types of response.

The experimental variable introduced by Patrick was emotionality. When the subject entered the room he received an electric shock on his bare feet, a horn blast in his ears, and a cold shower all over.

[3] Methodologically, it is noteworthy that this curve-fitting type of analysis requires a wide range of stimulus values on the independent variable, in this case numbers of dots from 1 to 210, and considerable precision in determination of the values on the dependent variable. The discontinuity would not be apparent in a narrow range of inaccurate data points.

Figure 2.3 shows the frequency of the five types of activity under calm and emotional conditions for five control subjects and four experimental subjects (the fifth experimental subject quit) ten times a day for ten days.

It is always hazardous, of course, to permit an observer who is aware of variations in the independent variable, in this case emotional excitement, to record and classify variations in the dependent variable. Nevertheless, it is quite clear from this experiment that the subjects displayed Type A activity most frequently when calm and Type E

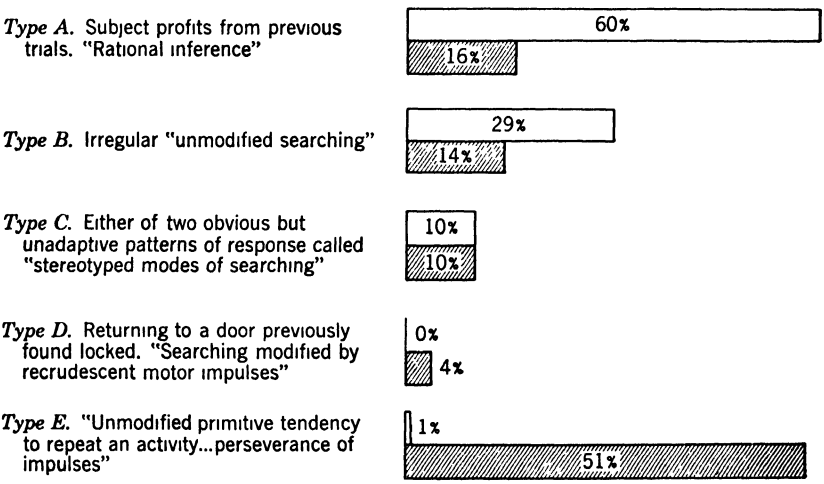


FIGURE 2.3. Effects of emotional excitement on five types of response. The experimeter classified 156 responses of five subjects under control conditions (open bars) and 198 responses of four subjects under experimental conditions (cross-hatched bars). Data from Patrick (*1934*).

activity when excited. The important methodological point is that valuable results were obtained from the combination of methods which would not have been obtained if either the subjective classification or the experimental variation of conditions had been used alone. Finding differences in frequency of types of activity under the two conditions validates the classification procedure.

A recent experiment by Maier (*1945*) suggests a quite different way of combining subjective and objective techniques advantageously. Maier had several groups of college students attempt his well-known

hatrack problem, which requires clamping two sticks together between floor and ceiling so that the clamp affords a solid place to hang a hat. Time is taken and success or failure is recorded quite objectively. In order to test some hypotheses about the processes by which the problem is attacked, Maier had other subjects watch him clamp the sticks together, then asked each to "write down all of the things or functions or uses he could figure out for the apparatus." A subjective classification of these subjective reports, combined with objective data from the other groups that actually attempted the problem, constitute the data from which conclusions are drawn. Specifically, since only a small percentage of the subjects attributed a hatrack function to the structure, Maier concludes that only rarely is the problem solved by immediate perception of an equivalence between the present situation and some experience of the past.

This conclusion will be discussed in a later chapter. At present the methodology is of particular interest and should be examined closely. Certainly the classification of the subjective reports would be more convincing if checked by an independent classifier. A more serious criticism is that the subjective data are taken at face value and really carry the burden of the conclusion. If the percentage of the subjects' responses classified as indicating a hatrack function are too low or too high, the conclusion is correspondingly strengthened or weakened. This way of combining objective and subjective procedures fails, therefore, to avoid the possible unreliability of the classification.

Patrick's procedure, by contrast, does not make such heavy demands on the reliability of classification, because he compared activities under experimental and control conditions, using the same classification procedure for both. Unreliability of classification would attenuate the chances of finding a significant difference, but would not systematically increase the chances of finding a spurious difference.

Regardless of the above criticism, Maier's ingenious use of the subjects' interpretations of the problem situation has methodological possibilities that should be exploited. The subjects could be asked to describe how they perceive a problem situation or certain aspects of it. These responses could perhaps be reliably classified, or the subjects might make their responses in multiple-choice form. A comparison of these subjective reports with type of solution attempt and with objective data on success and failure would probably help to clarify the role of perceptual processes in problem solving. If, for example, subjects who

report that they perceive a certain object in the problem situation as a screwdriver fail to solve the problem, while those who perceive it as a lever succeed, the reports of perception must have some reliability and perceptual organization must be important. Unreliability of report would yield inconclusive but not erroneous results.

The combination of objective and subjective data leads commonly to a methodological error that should be noted. Often objective experimentation shows that some subjects solved a laboratory problem while others did not, or that some solved it one way while others solved it another way, or that some made one decision while others made another. Significant differences are obtained. The experimenter then talks to the subjects in each group and gets their comments, which may be called introspections, about their intellectual operations. If the report of the experiment gives the subjective data, of unknown reliability, equal weight with the objective data, a hasty reader may accept the subjective distinctions as reliably established. Perhaps we have here an atmosphere effect: the test of significance performed on the objective data wafts an aura of respectability around the subjective analysis.

These illustrations point up the essential advantages of the combination of subjective and objective methods. Objective experimentation is convincing when clear-cut results are obtained, but the method may be too blunt to detect the subtleties of the thought processes. Subjective procedures are sensitive to the intangibles of thinking, but are often lacking in reliability and often distorted by the experimenter's or the subject's interest in the experiment and the latter's inability to communicate completely. If the research plan requires only a simple classification of the responses, not crucial in itself, reliability of classification may be obtained, and then the crucial evidence comes from more objective data, such as the frequency and time of responses in each category.

EVALUATION OF OBJECTIVE EXPERIMENTATION

The objective experiment is the most dependable method of psychology. As many thinkers have pointed out, however, some subjective or logical analysis must precede experimentation. It may be necessary, for instance, to identify responses that are to be counted as errors. It may be necessary to decide arbitrarily how to measure elapsed time or previous practice. But this is not a serious limitation, and it is certainly not peculiar to the investigation of the intellectual processes.

A more important logical consideration often overlooked is that the separation of two processes is not, *ipso facto,* the identification of either. Suppose, for example, that we wish to differentiate associative learning from reasoning or daydreaming or conditioning of Type X. The first requirement is a task that everyone will agree is an example of associative learning or some variety of behavior clearly identifiable as such. The other process must be equally well identified. If this can be done, the next task is to design an experiment in which the effects of some independent variable, e.g., practice, surgery, separate one from the other. If the two activities are not clearly identified, they may still be experimentally separated, but then argument begins as to what has been separated from what. The investigator can circumvent some of these difficulties in identification of familiar processes by inventing new ones, e.g., learning of Type Y, or reasoning of the Smith type. But he then has the task of relating these processes to others, of locating the new concepts in a system of concepts. If he neglects this larger task, his analysis remains a midget wisdom, isolated from the remainder of the science.

Judging from results to date, objective experimentation is not likely to furnish a system of thought processes as comprehensive as those that come out of subjective procedures, as described above, or factor analysis, to be described below. But when thought processes have been tentatively identified and systematized by either of these methods, objective experimentation is the method of choice for more precise description and analysis.

There is no difference in principle between an experiment on thought and judgment and one on perception or any other psychological function. In practice a higher degree of ingenuity may be required for analysis of the subtle processes involved and for the design of experiments that will illuminate the relations between them. Later pages will show that psychologists have had the ingenuity and that a good beginning has been made.

ANALYSIS OF INDIVIDUAL DIFFERENCES

It would be possible to discuss the thought processes at length and with profit solely in terms of general experimental psychology, ignoring individual differences. It would also be possible, though more difficult for historical reasons, to approach the topic from a consideration of individual differences in thinking without regard to experi-

mental information. In many textbooks and advanced treatises the discussion of differences in intellectual abilities makes practically no contact with the rest of the book. But, if psychology aims at a broad understanding of the topic, the separation of these two approaches is a tragedy. One of the major purposes of the present work is an integration of these two streams of progress. For convenience we may distinguish three techniques for analysis of individual differences, all of which have contributed valuable information to be reviewed in later chapters: comparison of groups, correlational analysis, and factor analysis.

COMPARISON OF GROUPS

Often, when a number of people have attemped a problem, the investigator has compared the successful with the unsuccessful in some respects. Or perhaps those who approached the problem one way or reached one decision have been compared with those who did something else. Comparisons have been made in respect to age, sex, education, experience, occupation, and the like, usually by a simple comparison of means.

The possible use of data obtained by subjective procedures should be mentioned. If it could be demonstrated, for example, that women spend more time than men on preparation, or that the successful thinkers report more revision than the unsuccessful, or that some occupational group used a particular attack more than some other occupational group, such information would be enlightening. The comparison of groups in this way is a familiar method which presents no special problems of experimentation or statistical treatment.

CORRELATIONAL ANALYSIS

The correlation technique enters the psychology of thought and judgment in several ways. If the problem is a complex one, the time of solution, number of responses, number of different responses, number of perseverative responses, and so on, may be correlated with other individual differences. Or, if many short problems are used, as in printed tests of arithmetic problem solving, a score representing the number right may be used for correlation purposes.

The correlation technique is applicable also to data obtained from subjective analysis. If a preparation period or a verification period can be reliably marked off and timed, such time scores could be correlated

with similar scores for the same thinkers working other problems and with scores for persistence, confidence, knowledge of logic, and other variables of background and personality. Significant relationships would argue for the validity of the method and would illuminate the problem-solving process.

Many technical difficulties arise in working with correlation coefficients, most of which are discussed in the more critical textbooks of statistics. One of the commonest statistical errors encountered in the literature is the neglect of attenuation. When scores from Test A are compared with scores from Test B and a correlation coefficient of .50 is obtained, the conclusion is often drawn that the two tests have little in common. But the correlation cannot be accurately interpreted unless the reliability of both scores is known—in the form in which the tests were used, and for the sample from which the correlation of .50 was obtained. The conventional correction for attenuation may have no practical value, but it is helpful in evaluating theoretical relationships between abilities. The disturbing fact is that when the reliability of a test is unknown, it is probably because the test has not been carefully prepared and refined, hence the reliability is likely to be low. The effects of attenuation must be considered not only when abilities are correlated but also when time scores and measures of persistence, confidence, perseveration, rigidity, or anything else are used for correlation purposes.

The correlation between two variables is also meaningless unless the variables are identified. Analysis of the thought processes is not aided by the knowledge that A and B are uncorrelated unless A and B have been studied, by subjective procedures, objective experimentation, or additional correlational analysis. The ideal of correlational analysis, as well as of the other methods, is not just the determination of the relationship between two variables but the construction of a system of relationships.

In addition to these complexities it is always difficult to infer the reason for the obtained relationships. If, for example, those who score high on arithmetic problems score high on spatial problems also, is the relationship due to a general ability, a speed factor, persistence in preparation, general good adjustment to schoolwork, or covariation in amount of exposure to both? Sometimes a combination of correlational and experimental methods assists the identification of the variables and the inference about causal relationship.

COMBINING CORRELATIONAL AND EXPERIMENTAL ANALYSIS. After Maier (*1929*) had separated reasoning from associative learning by a special kind of experiment, Campbell (*1935*) tested a group of rats on both tasks and found no correlation between the two. Similarly, English, Welborn, and Killian (*1934*), after separating substance memory from verbatim memory by a comparison of retention curves, showed that individual differences in the two correlate only in the .40's.

The rational and the analytical power of this combination of methods may be seen more clearly by returning to the separation by Kaufman *et al.* of subitizing from estimating (page 39). If the perception of small numbers of dots and large numbers of dots is performed by different processes, individual differences in subitizing would not be expected to correlate with individual differences in estimating. To test this inference a few correlations were computed from the data on which Figure 2.2 was based, as follows.

The discontinuity is in the neighborhood of 6 dots. Hence we may assume that the subitizing of 1, 2, 3, and 4 dots is different from the estimating of 7, 8, 9, and 10 dots. Correlations of individual differences within each of these two ranges would be high, but the correlation between the ranges would be low. Unfortunately, for present purposes only four subjects were used, but the data, median report times for 21 reports, are quite stable and the same subjects were used throughout the experiment. Transforming the median report times for these subjects subitizing 1, 2, 3, and 4 dots into ranks and computing the six rank-order correlations, we get a median *rho* of .50. The same procedure for 7, 8, 9, and 10 dots yields a median *rho* of .80. Hence the four "tests" of subitizing have much in common and so do the four "tests" of estimating. But when we correlate each of the four "tests" of subitizing with each of the four "tests" of estimating, we get sixteen cross-correlations, only two of which are positive, the median *rho* coming out as −.33. We must conclude, for this sample of four subjects at least, that correlational analysis confirms the analysis by curve fitting. Subitizing and estimating are different in two ways: they follow different laws, and they rank people in different orders.

In Chapter 9, which compares perception with judgment, these empirical results will be very helpful. Methodologically, it is noteworthy at this point that this empirical analysis yields a differentiation that would hardly have been intimated by logical or subjective analysis.

Most correlation studies have used raw scores representing time or number of successes. When the data have been described by a mathematical equation, the parameters of the equation, representing usually the slope of a curve and the theoretical upper limit, are available as measures of individual differences for correlation purposes. It is conceivable that such mathematically derived scores would contribute more to an analysis of interrelationships than raw scores which represent only one point on a curve. An application of this technique to the analysis of individual differences in the production of words will be described in Chapter 7, but whether this sophisticated procedure will actually be more fruitful than the use of cruder scores remains to be seen.

MULTIPLE CORRELATION. One practice of applied psychology that could well be imitated by general experimental psychology is the attempt to discover the most significant variables and to account for the variance on the dependent variable as completely as possible. Not infrequently an experimentalist, or even a school of them, becomes enamored of one variable or one process or one explanatory concept, and then devises new situations in which the favorite can be attractively displayed. In a well-controlled experiment variables which have no fascination for the experimenter are carefully kept out of the limelight. The psychology of thought has suffered from this tendency as much as any field of psychology, for it appears vulnerable to raids from learning theorists, psychoanalysts, and social psychologists, all eager to show off their research pets.

While such enthusiasm has definite stimulating value, an equally worth-while complementary enterprise would begin with the phenomenon to be accounted for, namely, the solution of a problem or the outcome of judgment, and attempt to predict individual variations in this phenomenon as completely as possible. The multiple correlation technique used in the selection of personnel offers a model for emulation in the laboratory. The result may be that the investigator's *a priori* analysis is confirmed, or that his favorite process has negligible predictive value as compared with others, or even that most of the variance on the dependent variable remains unaccounted for and a search for new significant processes is necessary.

The methodology of these prediction studies is well known. The point that needs emphasis is that the attempt to work out a multiple regression equation for the highest kind of intellectual activity would

force attention to the most important predictors of that activity. And just as the attempt to predict success in mechanical jobs has led to greater understanding of the ingredients of mechanical ability, a similar increase in understanding might be expected from an attempt to predict success in writing a poem, appraising real estate, solving crossword puzzles, or planning psychological research. Little information of this sort is available at present, but some beginnings will be noted in later chapters.

FACTOR ANALYSIS

The development of factor analysis has presented fascinating potentialities for the analysis of the thought processes along with difficult problems in technique and interpretation of results. People can be tested and scored on countless different variables, more or less related, but factor analysis promises to reduce this tangle of variables to a small manageable number and to clarify the relationships between them. Factor analysis is usually successful in that a sizable proportion of the variance of the scores on forty variables, for example, can be accounted for by the assumption of ten new statistically derived variables called factors. Whereas the relationships between the original variables are complicated and difficult to understand, the hypothetical primary factors are independent or at least arranged in a simple structure of relationships. The psychology of thought and judgment sadly lacks such an ideal system of relationships.

One difficulty in factor analysis is that a factorial solution of a set of correlations is not unique. One solution may reduce the forty variables to ten primary factors, while another may come out with six somewhat different factors. American factorists, for example, usually extract more factors from a matrix of correlations than British factorists. (For other British-American comparisons, see Vernon, *1950.*) The choice of solutions may lie outside the factor analysis itself, requiring illumination from further factorial investigations or new light from nonfactorial sources.

This lack of uniqueness is not peculiar to factor analysis. If our goal is a system of thought processes and the relationships between them, none of the methods of analysis described in this chapter promises an immediate unique solution. Systematization and clarification come through continuous research by different investigators, with different subjects and different tasks, and above all by different methods.

Furthermore, uniqueness may not always be desirable. For some purposes it may be desirable to identify factors that will be functions or processes; hence the test battery should call for a wide variety of operations on the test material, and the factors should have such names as perceiving, reasoning, memory, and attention. For other purposes it may be desirable to identify factors according to the kind of material involved, and the resulting factors might be such as verbal, spatial, and numerical.

The problem of most interest at this point is the interpretation and naming of the factors. The new factors, after rotation, are hypothetical mathematical constructs. The correlation of each factor with the original variables is known, however, so the interpretation begins with this knowledge. If we can infer, for example, that all the tests that correlate with a factor depend in some way upon knowing the meanings of words, and that none of the tests that do not depend upon word knowledge correlates to any extent with this factor, it is reasonable to call this new factor word knowledge. The procedure is to examine the tests that correlate with a factor in order to see what they all have in common and how they differ from those that do not correlate with this factor.

After the investigator has studied the tests that correlate with a factor and has formed a hypothesis of their common content, he may be able to construct a relatively pure test of that factor. If the test remains pure, that is, if it does not correlate with other factors in later factorial investigations, the processes tested by that test are more clearly identified. But the construction of a pure test of a factor is not always feasible. Some concepts remain hypothetical, like a complete vacuum, an absolute zero of temperature, and a free competitive market, yet serve useful scientific purposes. Factor analysis is able to handle a hypothetical factor, representing it in factorial space and calculating its correlation with existing tests, even though a pure test of the factor cannot actually be constructed.

There is no guarantee, *a priori,* that every activity can be represented by a pure test or even by a linear combination of scores on independent tests. It is possible that some complex intellectual activities do not entirely conform to the assumptions of factor analysis. As a matter of fact, most factor analyses to date have left much of the variance of the original scores unaccounted for. Whether this is a permanent limitation of the method or a temporary characteristic is

not known. And the other methods do not offer completeness of analysis either. Multiple correlation hardly ever accounts for more than half the variance on the dependent variable. Subjective analysis lacks the reliability to permit the necessary calculations. And experimental analysis does not even try to be complete.

The description of the abilities involved in thinking may present special difficulties for factor analysis. Factor analysts like to work with pure tests, representing simple abilities. But thinking is not simple. Many intellectual operations occur simultaneously. Though it may prove impossible to fractionate thinking into independent factors and construct a pure test of each factor, the abilities involved in thinking may be described by including complex thought problems in the battery subjected to factor analysis. Rotation can ignore the complex thought problems, but after the factors are identified the correlation of each hypothetical factor with problem-solving success can be computed. In this way the abilities involved in complex unanalyzable intellectual activities can be described in reference to simple independent factors.

Since factor analysis is a recent invention and since the naming of the rotated factors is a subjective procedure, the new factors have usually been described in respect to categories that the older methods had made familiar. The reverse may someday occur. Variables isolated by factor analysis may later be analyzed and further described by experimental methods.

There are many technical difficulties in the way of factor analysis that need not be considered here. Some of these are being attacked, and perhaps it may be said that there is nothing wrong with factor analysis that better factor analysis will not cure. But the sheer labor involved indicates that this is no easy way to analyze intellectual activities. Factor analyses based on less than ten variables or less than a hundred subjects are hardly worth the trouble. Because of the youth of the method, the results to be presented in later chapters are not extensive, but they do shed considerable light on our topic.

TRIFUNCTIONAL ANALYSIS OF THOUGHT PROCESSES

Each of the methods discussed in this chapter has contributed, in its own way, to present knowledge of the thought processes. At this point our need is for a set of categories representing thought processes to

be used as chapter headings for a systematic survey of the available information. In this hour of need, logical analysis is our best friend; it gives the broad outlines of the field within which more objective evidence can be arrayed.

Certainly one chapter must deal with the preparation for intellectual activity. Logical analysis requires it. Subjective analysis makes the concept familiar, and Wallas specifically emphasized a preparation period by name. In the simpler experiments a preparation period has been objectively marked off, timed, and described in neuromuscular terms.

Following preparation, there will be two chapters on the flow of ideas or productive thought. Classical introspection reported that ideas "come to mind" or "appear in consciousness." William James made every reader conscious of his own "stream of consciousness." Most other subjective analyses have included this production process in one way or another; some investigators have tried to record thoughts as they are produced. Experimental psychology has considerable objective information on the production of simple responses, such as words and movements, or sequences of these, and even on the production of large abstract patterns. In several factors analyses fluency of production of words and ideas has been identified as a factor of some importance.

The third process, judgment, may be identified as the evaluation or categorizing of an object of thought. This is logically differentiated from productive thought in that typically nothing is produced. The material is merely judged, i.e., put into one category or another. Many of the subjective analyses of thinking have included a concluding phase of hypothesis testing or verification during which the thoughts previously produced are judged. In experimental psychology, judgment is a well-developed topic, studied chiefly under the headings of psychophysics, aesthetics, attitudes, and ratings of personnel. Analysis of correlations between individual differences in judgment is a recent trend. At least one large-scale factor analysis has tentatively identified a factor called judgment.

These, then, are the principal thought processes which fix the main chapter headings for this book: preparation, production, judgment. We shall use terms like thought, intellectual activity, and problem as very general terms, to be qualified as appropriate. Judgment is a special kind of thought. A judgment problem is a special kind of

problem. Armchair analysis and synthesis of published research establish the gross outlines or topical organization; within this framework the evidence will be more objective, coming from experiment and analysis of individual differences. The fact that this triphasic analysis is convenient for a review of the psychological literature gives it a sort of literary validity. A more critical analysis will be attempted at the beginning of each chapter.

LEVELS OF ANALYSIS. As soon as these categories are set down, the question of scale or level of analysis immediately arises, both to plague the analysis and to remedy its ailments. To settle this question let us examine three typical thought problems, of increasing complexity.

"What is the opposite of *congenial?*" Subjectively our three processes seem to describe what goes on quite well. The thinker ceases what he had been doing, brings his attention to bear on this problem, and perhaps reviews the meaning of *congenial* (preparation). Then words are produced, one after another (production). These words are serially evaluated and one is accepted (judgment). So far so good. It will require at least three chapters to describe the solution of a problem, but our three processes are coördinate categories which account for the activity rather completely and can be separated from each other in time and in function.

"What is the opposite of *square?*" The thinker may prepare as before and produce words like *round* and *triangular,* then decide that none of these adjectives is really the opposite of square. So he turns back and questions his initial preparation. "Are there other meanings of *square?* Could it have an ethical instead of a geometrical meaning? Should it be considered a noun or a verb?" Here we have production subsidiary to preparation. One may produce several ways of setting up the requirements of the problem before undertaking the actual production of solutions to the problem. And if our thinker produces several alternative preparations of approximately equal attractiveness, he will have to choose among them tentatively at least before getting into production. Hence judgment also may be subsidiary to preparation.

In the same way production may be subsidiary to judgment. Suppose the thinker takes *square* as a verb of mathematical meaning. He might then produce *extract the square root of* as an opposite. Then when he judges the correctness of his solution, he may ask himself whether there are other angles to the problem that he has overlooked.

He then produces all the angles, considerations, logical fallacies, and relevant memories that might aid the evaluation of his solution.

We must admit that the solution to a relatively simple problem is not a straightforward sequence of events. The operations are geared together as a dynamic meshwork in which any may be activated by, and function subordinately to, the others. This triphasic interaction is what gives thought its familiar wheels-within-wheels character.

For our third problem let us consider the case of a man trying to design a house. The sequence of processes will be so complicated that it is hardly worth while to try to follow it. No doubt the overall problem will be broken up into subproblems, and one of these subproblems may lend itself to analysis, but little can be proved about the overall problem by analysis of sequences. Thoughtful solution of a problem is a limited episode in a busy life. It must begin with some sort of preparation and end with some sort of conclusion or decision. But between these limits anything can happen in any sequence.

RESEARCH STRATEGY. The psychology of thought cannot ignore large lifelike problems like designing a house, even though they are hard to analyze. The usual procedure of science is to assume that the operations identified in simple instances also occur in complex instances, and to arrange empirical tests of the consequences of these assumptions. Problems can be devised with various degrees of emphasis on any one process, the other processes being held constant.

In any complex act of thought any of the three processes may be critical; difficulties may be located anywhere, and any process may be the source of individual differences in the final outcome. But there are standard psychological tasks that emphasize one of these processes by minimizing the role of the others. In some experiments on reaction time and controlled association, modification of the preparation has been the principal independent variable. Experiments on word fluency emphasize the production of words with no special preparation and no judgment of the output. Psychophysical experiments require no productive thought, the material coming directly from sensory processes, hence the judging process is primary.

The research strategy in the experimental isolation of the thought processes is analogous to the experimental isolation of the three phases of learning: acquisition, retention, and recall. The invention of methods for controlled observation of the operation of the three phases of learning is one of the triumphs of modern experimental psychology

and could well serve as a model for investigation of thinking. And just as the investigation of learning must begin with some preliminary considerations, such as the role of motivation, the investigation of thinking must devote a few chapters to preliminaries before getting into the thought processes proper.

REFERENCES

Allport, G. W., and Odbert, H. S., 1936. *Trait-names: a psycho-lexical study,* Psychol. Monogr., No. 211.

Bloom, B. S., and Broder, L. J., 1950. *Problem-solving processes of college students, Suppl. educ. Monogr.,* No. 73, Chicago, University of Chicago Press.

Blumenfeld, W., 1931. Urteil und Beurteilung, *Arch. f. d. ges. Psychol.,* Ergänzungsband 3.

Burack, B., 1950. *The nature and efficacy of methods of attack on reasoning problems,* Psychol. Monogr., No. 313.

Campbell, A. A., 1935. Community of function in the performance of rats in alley mazes and the Maier reasoning apparatus, *J. Comp. Psychol., 19:*69–76.

Cattell, R. B., 1943. The description of personality: 2. Basic traits resolved into clusters, *J. Abnorm. & Social Psychol., 38:*476–507.

Claparède, E., 1934. La genèse de l'hypothèse, *Arch. de Psychol., 24:*1–154.

Dewey, J., 1910. *How we think,* Boston, D. C. Heath & Co.

Durkin, H. E., 1937. *Trial-and-error, gradual analysis, and sudden reorganization,* Arch. Psychol., No. 210.

English, H. B., Welborn, E. L., and Killian, C. D., 1934. Studies in substance memorization, *J. Gen. Psychol., 11:*233–260.

Guetzkow, H., 1950. Unitizing and categorizing problems in coding qualitative data, *J. Clin. Psychol., 6:*47–58.

Harlow, H. F., 1949. The formation of learning sets, *Psychol. Rev., 56:*51–65.

Heyns, R. W., 1948. Functional analysis of group problem-solving behavior, University of Michigan Conference Research (unpublished).

Jensen, E. M., Reese, E. P., and Reese, T. W., 1950. The subitizing and counting of visually presented fields of dots, *J. Psychol., 30:*363–392.

Kaufman, E. L., Lord, M. W., Reese, T. W., and Volkmann, J., 1949. The discrimination of visual number, *Am. J. Psychol., 62:*498–525.

Leeper, R., 1951. Cognitive processes, in *Handbook of experimental psychology,* Ed. S. S. Stevens, 730–757, New York, John Wiley & Sons.

Lowes, J. L., 1930. *Road to Zanadu,* Boston, Houghton Mifflin Co.

McClelland, D. C., and Apicella, F. S., 1945. A functional classification of

verbal reactions to experimentally induced failure, *J. Abnorm. & Social Psychol.*, *40*:376–390.

Maier, N. R. F., 1929. *Reasoning in white rats,* Comp. Psychol. Monogr., No. 6.

Maier, N. R. F., 1945. Reasoning in humans, III. The mechanisms of equivalent stimuli and of reasoning *J. Exper. Psychol.*, *35*:349–360.

Patrick, C., 1935. *Creative thought in poets,* Arch. Psychol., No. 178.

Patrick, C., 1937. Creative thought in artists, *J. Psychol.*, *4*:35–73.

Patrick, J. R., 1934. Studies in rational behavior and emotional excitement, *J. Comp. Psychol.*, *18*:1–22, 153–195.

Platt, W., and Baker, R., 1931. The relation of the scientific hunch to research, *J. Chem. Educ.*, *8*:1969–2002.

Reed, H. B., 1946. The learning and retention of concepts, *J. Exper. Psychol.*, *36*:166–179.

Reitz, W., 1950. The higher mental processes, in *Encyclopedia of Educational Research,* ed. W. S. Monroe, 540–551, New York, The Macmillan Co.

Rossman, J., 1931. *Psychology of the inventor,* Washington, D.C., Inventors Publishing Company.

Symonds, P. M., 1936. *Education and the psychology of thinking,* New York, McGraw-Hill Book Co.

Taves, E. H., 1941. *Two mechanisms for the perception of visual numerousness,* Arch. Psychol., *37*, No. 265.

Thurstone, L. L., 1930. The learning function, *J. Gen. Psychol.*, *3*:469–493.

Thurstone, L. L., 1948. Psychological implications of factor analysis, *Am. Psychol.*, *3*:402–408.

Vernon, P. E., 1950. *The structure of human abilities,* New York, John Wiley & Sons.

Wallas, G., 1926. *The art of thought,* New York, Harcourt, Brace & Co.

Chapter 3

The Dynamics of Thought

> Man began to think in order that he might eat, now he has evolved to the point where he eats in order that he may think.
>
> —W. P. Montague

IT would be possible to analyze the behavior of organisms in their environments, including the human organism in a problematic environment, without invoking dynamic principles, but the program would be unnecessarily long and inconvenient. The activities of man and the other animals are more comprehensive, more amenable to control and prediction, when the internal state of the organism is taken into account. For this reason modern psychology speaks often of motivation, set, and other dynamic concepts. These hypothetical internal engines and gears, presumed to energize and control externally observable behavior, are promising explanatory constructs. They should not be shoved hastily and illicitly into the theoretical apparatus when it fails to work, but their functions and relationships should be built into the system and clearly described at the outset. This is best done by proceeding from the outside inward, beginning with motives, which are quite general, and working toward sets, which are more specific.

MOTIVES FOR THOUGHT

It is not necessary here to settle all the troublesome questions of the psychology of motivation. It is necessary merely to describe the kinds of motives that influence thought and judgment.

PRIMARY MOTIVES. In modern psychology the principles of motiva-

tion have usually been constructed after the model of biological motivation or animal drive. Hunger, thirst, and sex are taken as typical. Physiological changes integrate the organism so that energy level is raised; perception is sensitized to certain promising patterns of stimulation, in human beings even to the point of fantasy; and the effector apparatus is prepared for the goal response. Typically the goal response has survival value either for the organism or the species. Some reflexes, such as withdrawal from pain and emptying of the bladder, may, by involving the whole organism when the usual segmental response is prevented, fit into this scheme of biological motivation. Hunger and punishment are commonly used to motivate animals in problem-solving experiments. Punishment is often part of the pattern of motivation used with human subjects.

Social Motives. Much human behavior is directed toward people or activities involving people. So the same formulation of motivational principles is applied to social behavior. A person who has been deprived of companionship is more sensitive to patterns of stimuli that promise companionship, even to the point of fantasy, is more ready for social intercourse, and will do more to attain it than the same person when he is surfeited with people. Under certain conditions tremendous drive is directed toward moving upward in the social hierarchy. The person becomes sensitized to opportunities for attaining prestige, and may even manufacture dreams of glory before his goal is reached.

Social motives connected with the concept of the self are often called ego motives since this self-concept or ego is a central value which most people will vigorously defend in recognizable ways, e.g., by rationalization. The goal is self-respect, a satisfactory picture of oneself. In some circumstances great energy is released and directed to this end, and much thought is daily devoted to ego problems. Furthermore the results of thought, the gadget invented, the poem created, the decision reached, any kind of brain child, may become a part of the ego, to be defended as strenuously as one's face or pocketbook. Thus the degree of ego involvement in intellectual matters becomes an important dynamic consideration for thought and judgment.

These social motives differ from the primary biological motives in that there is no special physiological basis for the former. The integration of behavior is a learned one, and the explanation is developmental rather than physiological. The events of the individual's life history, such as social stimulation in childhood, success and failure in social

relations, and the expectations of family and culture, are known to be influential. Since the social motives are presumed to be acquired during the operation of the primary motives, they are often called derived motives.

Since human beings are very plastic, able to develop in many directions, the vicissitudes of life on this earth foster the acquisition of many other secondary motivational trends. Almost anything that can be identified can become a goal to be attained or an annoyance to be avoided. No competent psychologist has claimed to understand the nature and origin of all motives. Whatever the origins may be, no one will deny the attraction of such goals as prestige, the praise of one's friends, and self-respect in motivating the solution of life's and the laboratory's problems. The motivation for most laboratory experiments on problem solving and judgment is largely social.

It is important to note, in respect to both classes of motives, that this conventional formulation of motivational principles separates the motivation from the activity motivated, the drive from the mechanism. Under the influence of the hunger drive a dog may learn the location of a dish, the manipulation of a latch, or skill in standing on his hind legs. And, in reverse, any one task, e.g., standing on the hind legs, may be mastered under the influence of hunger, thirst, or social motivation. The constant features of any motive, according to this formulation, are the internal motivating conditions and the goal response; between these two activity is varied. Here is where problem solving and learning come in. If the motivated organism cannot proceed directly to the goal, it is said to be in a problem situation. New patterns of activity are developed and practiced as ways of attaining the goal. The motivation of activity is originally extrinsic to the activity itself; rewarded practice weaves the two into a pattern of adjustment to the exigencies of life.

Intrinsic Motivation. If the familiar statement that "all behavior is motivated" means that all behavior is caused, no one with any faith in science will object. But if it means that all behavior is extrinsically motivated, exceptions can be noted, some of which are particularly pertinent to the dynamics of problem solving. No one, for instance, who has watched children, or adults, at an amusement park, can convincingly maintain that the goings-on observed there are motivated by some extrinsic biological or social motive. A ride on a frightening roller coaster is fun. Its attractiveness is an end in itself, not a by-

product of some remote primary motivation. When a healthy dog is let out on a frosty morning, there seems to be no motive *behind* his energetic cavorting. In these and similar instances, separating the motive from the activity is as artificial as separating the dog's muscles from his muscular contractions. Observation of young children's noise-making, block-building, and digging activities will yield many other examples of apparently self-motivated behavior. It is necessary to admit that some sensory stimuli and some activities have positive and negative motivation value *for their own sakes,* just as a lump of modeling clay, by its own intrinsic charm and plasticity, invites manipulation. Since these activities are performed without the spur of practical or extrinsic motivation, they are often listed under the heading of play.

Activities listed under the heading of curiosity or exploratory drive illustrate the same sort of intrinsic dynamics. Most animals remain active for a while in an unfamiliar environment until they have explored their surroundings, noting openings and obstacles and, in general, getting the lay of the land. They all register the illumination and adjust their stance to the inclination of the substratum quickly and automatically. Rats will cross an electrified floor and take considerable shock in order to get to the other side and nose around in paper and sawdust (Warden, *1931*). They will spend more time exploring a novel stimulus than a familiar one (Berlyne, *1950a*).

The survival value of a survey of the potentialities and limitations of the immediate situation is obvious. It is unfortunate that the experimental investigation of animal motivation has been so largely limited to motives that are easily aroused and controlled in the laboratory. Careful observation of animals in a wide variety of environments, doing the things they naturally do, might broaden the psychology of motivation.

Monkeys seem to be naturally curious, fond of playing with movable objects. Harlow, Harlow, and Meyer (*1950*) have demonstrated that rhesus monkeys will take a puzzle apart when they find it in their living quarters even though not extrinsically rewarded for doing so. Solution of the puzzle, sketched in Figure 3.1a, required three acts in correct sequence. First a pin on the end of a chain had to be pulled out of a hole. This permitted a hook to be moved to one side. With the hook out of the way a hasp could be raised. No reward was given for solution, and social motivation was minimized.

Four monkeys had the assembled puzzle to play with for twelve days. As a control, four matched monkeys had an unassembled puzzle similarly available. When tested on the thirteenth and fourteenth days, the experimental monkeys were much more successful in all respects.

Finally a food reward was placed under the hasp to test the influence of extrinsic motivation. But this definitely disrupted performance. The monkeys always attacked the hasp directly, neglecting the previously learned sequence of movements. Evidently the motivation for solution had been an intrinsic one; the experimenters called it a "manipulation drive." But the activity initiated by a manipulation drive may later be kept going by other drives. When a group rewarded by food and a no-food group were compared, the no-food group was superior at first, but after two weeks and several thousand openings the food-reward group became relatively more active (Davis, Settlage, and Harlow, *1950*). Even more striking results have been obtained with the six-device puzzle shown in Figure 3.1b (Harlow, *1950*).

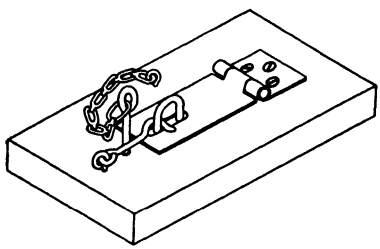


FIGURE 3.1a. Puzzle device used by Harlow, Harlow, and Meyer (*1950*). Rhesus monkeys solved this puzzle without extrinsic motivation.

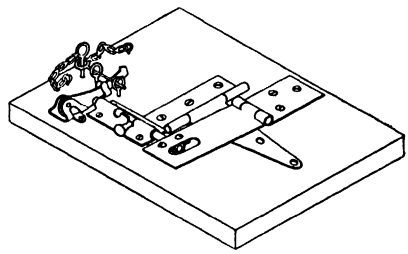


FIGURE 3.1b. Six-device puzzle used by Harlow (*1950*).

If monkeys will manipulate a puzzle just for the fun of it, human beings would be expected to do the same. People do work crossword puzzles, jigsaw puzzles, and many other kinds just to "occupy their minds." Attempts to explain away such mental exercise in terms of social motivation fail if the puzzler works by himself and does not communicate his successes—as many of them do.

A good illustration of the human love of puzzles occurs regularly

in psychology classes. The instructor introduces a puzzle to the class in order to clarify a point in his lecture on problem solving. Many of the students become absorbed in the puzzle, spending the remainder of the hour working on it, completely missing the principle that the instructor was trying to illustrate, and risking loss of the extrinsic rewards administered by the instructor.

The tendency to complete an unfinished task may be listed as intrinsic motivation. If someone is asked to solve an interesting problem, then, as he nears the end, is asked to stop, he may be so involved that he finishes it anyway. The task itself captures and directs his efforts after the extrinsic social motivation is turned off.

Perhaps intrinsic motives are weak compared with such strong motives as hunger and rivalry. But applied at the critical point, as they often are, their effect recalls the fable of the sun and the wind.

Claiming that an activity is intrinsically motivated is not the same as claiming that the motivation is innate. Evidence for or against inheritance of a manipulation drive or an interest in number games will be just as hard to produce as evidence on the inheritance of any other psychological trait. Allport (*1937*) has given many instances of, and a strong argument for, functional autonomy of motives without denying that a motive functionally intrinsic today may have some old developmental connection with a primary drive.

But the separation of intrinsic and extrinsic motivation of activity is really just a matter of emphasis. Although some motives are set going primarily by internal physiological changes, the external situation cannot be entirely left out of the picture. The hungry or ambitious organism may go looking for appropriate goal objects, but it is more likely to do so in some stimulus fields than in others. If the state of the organism is primary, the dynamics may be crudely pictured thus: O–s–R. On the other hand, the intrinsic appeal of a puzzle may be primary, but curiosity will be aroused more intensely when an organism is alert than when it is fatigued. In the Minnesota experiments on acute vitamin-B deficiency in normal young men, the decrease in self-initiated mental activity was marked (Guetzkow and Brozek, *1946*). The general responsiveness of the organism, its readiness to adjust to the environment, must never be overlooked. Given this general readiness to act, an intrinsic motive inhering in a stimulating object or activity may supply the more specific direction of be-

havior. This emphasis may be pictured thus: S–o–R. And of course intermediate degrees of emphasis occur as well.

PATTERNS OF MOTIVATION

Everyone knows that motives do not operate in isolation. It is an organism that is aroused, not a hunger drive or a prestige drive. Treating motives one by one is a tremendous oversimplification that may be necessary for exposition and experimentation, but the fractionated organism should be put back together again as soon as possible.

Typically, several motives are in operation simultaneously, behavior being initiated and directed by a motivational pattern or system of forces. These concurrent motives may reinforce each other or they may conflict. Thus it happens that objects and activities are sought as goals, or avoided, because of the simultaneous influence of several motives, some or all of which may be unrecognized. What is recognized is the attractiveness of the goal. That is why it is much easier for a person to say what he likes and dislikes than to say what motive lies behind these facts. An animal's activity in relation to objects and places in the environment may be observed even though the motives influencing such activity are not easily inferred.

Anyone who likes puzzles may entertain himself by attempting an analysis of the motivation of artistic activity, for this is a puzzle that psychologists have largely ignored. Surely social motives are prominent in some cases, but they can hardly carry the effort very far. The artist who keeps at his work in spite of personal privations and lack of recognition is probably motivated by a love of the medium for its own sake, enjoying the arrangement of sounds, the manipulation of pigments, the creation of appearances, with an intrinsic perceptuo-motor affection. An ego drive of significance for creative activity, and probably also for less exalted activities, is self-expression, the urge to extend the self by putting part of one's inner life on display, perhaps even clarifying one's relation to the rest of the world. Even these intangible and questionable motives may not tell the whole story; for when the going gets tough, many motives are extinguished. The artist may be helped along at this later state by the pressure for completion, the intellectual abhorrence of a gap or imbalance in the pattern.

Although the relative importance of these different motives cannot at present be estimated with any accuracy and the structure in which they are integrated cannot be described, we must remember the variety

and flux and organization of motives as we attempt to comprehend the details of intellectual activity. This volume is concerned with the most sophisticated, abstract, civilized, and specialized things that people do. A one-sided dynamic would certainly lead us astray.

WHAT IS A PROBLEM?

One of the reasons for the above survey of motivation is to introduce the concept of problem, for no one can have a problem without being motivated. *A person may be said to have a problem if he is motivated toward a goal and his first goal-directed response is unrewarding.* Typically he varies his activity, and perhaps by some subsequent response he does attain his goal or progress toward it. Problem solving, then, which is one meaning of thinking, is the activity, by definition variable, of an organism trying to attain a not easily attained goal. Stated thus, a problem is a personal thing. It is not a characteristic of a situation. It is had by a person, or an animal, in a certain situation, and to call a situation a problem situation is to assume that the subject of the discussion is motivated toward some goal in the situation that he will not attain on the first try. The activity is not the problem; to call a person's activity problem solving or thoughtful activity is to assume that such activity is varied in an attempt to reach the goal. Some psychologists, attempting to avoid words like "try" and "attempt," characterize a problem situation as one for which the organism has no appropriate or adequate response. But obviously such words as "appropriate" refer to motivation toward a goal. A response cannot be appropriate without being appropriate to some course of action. If one is not willing to assume that the organism's behavior is goal-directed, he cannot approach the essence of problem solving but must limit himself to superficies.

The key words in the italicized statement above are "first" and "unrewarding." The concept of problem solving refers to a sequence of activities because, if the motivated organism attains its goal on the first try, even though it be a lengthy try, it has no problem. The man headed for home on a familiar road has no problem despite the time required before reaching his goal. His first step is effective, and so is his second, his third, and so on. The man with the problem is the one whose first step, as he perceives it, is not effective, not rewarding.

The word "rewarding" also needs clarification. To say that a response is rewarding is to assume that the organism is motivated and

can discriminate between reward and no reward. In many laboratory experiments this discrimination is easy. The food-motivated rat jumps in the most promising direction and bumps his nose. If the motivation remains strong, he clearly has a problem. His first response was unrewarded and he will vary it. A man reaches in his pocket for his keys. If his hand closes on them with a rewarding thrill of familiarity, his progress is uninterrupted, and he clearly has no problem. If the pocket is empty, his first response is unrewarded: he has to stop and think. But suppose the man in the paragraph above is walking home free from problems when he notices something unusual about the trees. Has he a problem? Since his walking does not reward or punish him immediately, as the rat's jumps do, a separate act of judgment occurs. If he decides that the unfamiliarity is illusory and his activity is really furthering his homeward progress, he has no problem, no need to think. But if he decides that his course is unrewarding, is not taking him home, a thoughtful episode begins. In the complicated situations of real life, motivation is complex; rewards and punishments are often delayed. The effective motivation may not be all-or-none; it may be a gradient or pattern, some parts of which are clearer than others. In such dynamically complex situations the definition of a problem may require a separate act of judgment.

SET FOR SINGLE STIMULUS AND RESPONSE

As a rule a person's activities are integrated. He does not react to every passing event that stimulates his sense organs, and does not obey every biochemical whimsy. His perception is selective and his activities are more or less continuous in spite of environmental and visceral changes. This is one of the most distinctive properties of organismic behavior, and to study it thoroughly would be to study life itself; but the study of the integration or control of activity over a short time span is altogether feasible, and has in fact been a favorite of the psychological laboratory under such names as attention and set (Dashiell, *1940;* Gibson, *1941;* Paschal, *1941*).

Set is a particularly slippery concept. It has been used unsystematically to refer to a variety of phenomena, hence its psychological meaning is not clear-cut. Some psychologists have even raised the question whether the term is an asset or a liability to psychology. Actually there is a need in psychological dynamics for a concept like set, and, if we throw out set, we would have to construct a new concept to do

its job. The new concept could be more rigorously defined, but might then be out of contact with the wide range of research data now available. If the terminological confusion could be put right by throwing out a word, the improvement would be worth the price. But such a remedy would give only temporary relief; a new system of concepts would soon be called for. Hence, there is more to be gained at the present moment by accepting the present concept of set, defining it as clearly as possible, and showing where it fits into the larger system of concepts. A sophisticated use of the concept and a strong argument for its utility have been presented by Woodworth (*1937*) under the name of situation-and-goal set. A systematic survey of the evidence now available on set will show that this concept does help to clarify the dynamics of thought and judgment.

If motives supply the general long-term policies and goals of life, sets operate within this general dynamic as transmission mechanisms, geared to immediate stimuli and responses. A set may be defined, in the usual way, as a readiness to make a specified response to a specified stimulus. This is as good as most definitions, but its implications require some discussion. The most important characteristic of set, which does not appear explicitly in the definition, is its unity and continuity. The individual is subject to many demands, both personal and environmental, but at any moment behavior is unified or integrated. Environmental stimulus patterns are constantly changing, especially for the higher animals, which are mobile, and especially for man, who manipulates his environment. There are many motives and many exciting stimuli, but at any moment only one set. The term readiness in the definition means that the stimulus-response coördination is prepared in advance, so that when the stimulus is perceived, the response follows with little delay. On the stimulus side, the organism is prepared to select or abstract from the complexities of the environment certain stimulus patterns rather than others. On the response side, the organism is prepared to select from its repertoire of responses certain acts rather than others. Set has an implicit negative meaning also, since to be set for one kind of stimulus necessitates the neglect of others and to be set for one kind of response necessitates the inhibition of others.

Getting set for a stimulus implies the ability to identify the stimulus. If the stimulus is an abstraction, this implies the ability to abstract. One cannot get set to underline hostile words unless he can

abstract hostile words from others. Errors may result either from difficulty in abstraction or from difficulty in maintaining the set. This distinction will be important in later chapters where abilities are discussed.

In any workable system of concepts, set must function at an intermediate level of integration, between the overall motivational pattern and the specific little acts. If, over a period of months, a person reacts positively to such symbols as free speech, democracy, and free enterprise, the integration behind his reactions should not be called a set but an attitude, sentiment, interest, or value. If, however, he is looking through the evening paper to underline such words, he has a short-term integration or set, one that will disintegrate when his job is finished. For purposes of this book at least, set is a temporary integration of behavior that lasts longer than the reaction time but not so long as a motive.

Granting, then, that set is a concept of intermediate level, functioning between motives and acts, there is still much variation among sets. A person can be set for obvious things like a word or a danger signal, or for subtle aspects of complex stimulus patterns, as when looking through a manuscript for grammatical errors or rating people in respect to sociability. A person can be ready for a simple response, like pressing a key, saying a word, or putting a check mark on a dotted line; or for a complex sequence of movements, like starting an automobile when the light changes, or playing a musical instrument when the conductor gives the signal. The set may be quite specific on the response side, as in the case of the muscular set of the reaction-time experiments or of the sprinter at the starting line. Or the set may be quite specific on the perceptual side, as in the case of the sensorial set of the reaction-time experiments. A spectator may be keenly watching to see which runner breaks the tape first without being prepared to do anything about it. A person can be prepared in advance for any stimulus or any aspect of a complex stimulus situation that he can identify and for any activity that he can perform.

If this concept of set seems too general, too vague to be scientifically useful, let us raise the question of predictability. In the reaction-time experiment, if we know what instructions were given to the subject, we can predict his reaction better than if we lack this information. Certainly we cannot predict his behavior very well if we do not know how he has been prepared. If someone is observing children at play,

we cannot predict his judgments at all unless we know whether he is observing them in respect to sociability, general activity, or emotionality. Parenthetically, this is one of the advantages of human over animal subjects for controlled experimentation. Human subjects can be prepared by verbal instructions to observe almost anything and to make almost any response.

Then let us ask, in order to make the concept clearer, under what conditions the set does not aid in predicting behavior. When our subject hears a sudden noise or is struck an unexpected blow, his behavior will be hard to predict. In other words, when a stimulus is perceived for which the subject is not set, we cannot predict what he will do.

Thus defined, this three-letter word, set, carries a heavy conceptual burden. The argument for recognizing set as a useful psychological concept, and the parallel necessity for including set in a systematic treatment of thought and judgment, are best brought out by examining the conditions that establish a set and the effects of the set in operation, in short, by tying the concept down to the facts at both ends. Later, evidence on individual differences in set will confirm the evidence from experimental variation of the conditions of set.

HOW SETS ARE ESTABLISHED

Sets, as defined above, are established by (1) motives and (2) social interaction with other people. In addition (3) a set previously established may, like any other learned activity, be reinstated on a later occasion. Furthermore, a new or *ad hoc* set (4) may be organized during the performance of a complex task.

A Set May Be Established by a Motive. Hungry animals are more sensitive to food than well-fed animals and are prepared for the goal response. The perceptual side of the set for food has been demonstrated in human subjects by Sanford (*1936, 1937*); Levine, Chein, and Murphy (*1942*); and McClelland and Atkinson (*1948*). Sanford had children look at ambiguous pictures before and after meals and report what they saw. Food responses were twice as frequent before meals as after meals. It is important for later application of this principle to note that the effect is negative as well as positive—in accordance with the definition of set. Experiments with words as stimuli show that the perceiver may inhibit perception of nasty words (McGinnies, *1949*).

The same readiness can be observed under the influence of social motivation. When two people are competing intensely, each is quick

to note an opportunity to get ahead of the other and quick to carry it out. The drive for achievement influences perception of objects in an ambiguous stimulus field approximately as the hunger drive does (McClelland, Atkinson, and Clark, *1949*).

Certainly motive and set overlap. In the noncommunicative animals there would be no necessity to separate the two concepts. But human beings are more complicated than other animals; to understand them requires more concepts. Sets are established by motives, but the two concepts are not identical because in human beings sets are established in other ways.

A Set May Be Established by Social Interaction. Since people are sociable and communicative, they frequently prepare themselves to perceive certain stimuli and to execute certain responses simply because someone asked them to. The effect on human activity of instructions from other human beings is a remarkable phenomenon that should never be overlooked. Perceptual sensitivity to food stimuli and a readiness to eat can be prepared in coöperative human subjects not only by hunger but also by a request to act hungry. In fact, instruction of the subject by the experimenter is the standard procedure in the laboratory for establishing a set. The subject is instructed to watch for a light in a certain location and to press a key as soon as he sees it. Or he is instructed to say "right" if the light on the right is brighter than the light on the left. People can be prepared by instructions to hear faint sounds, to see faces hidden in drawings of scenery, to note typographical errors, to perceive almost any describable stimulus pattern and to make almost any responses within the limits of their capacities. They can get set in advance to give a response word that is a synonym or an antonym of the stimulus word. A large factor in the achievements of experimental psychology is just this ability of human beings to prepare their receptive and reactive functions when requested to do so.

The effects of social interaction are not limited to explicit instructions. Some social influences, such as suggestion, establish sets in an indirect fashion. When anyone is trying to solve a mechanical puzzle, for example, his activities may be slanted toward a certain place or a certain type of manipulation by comments of bystanders or misinterpretation of an observer's words or gestures. Prestige suggestion, as we shall see later, often exerts a strong influence on the set that people adopt when they judge a complex object, leading to an emphasis on

some aspects of the object of judgment and neglect of other aspects. In the laboratory, if the experimenter seems to be in a hurry, the subject may speed up. In the clinic, interpersonal relations between examiner and subject slant responses to the Rorschach test (Gibby, *1951*) and to a sentence completion test (Meltzoff, *1951*). Occasionally the effects of suggestion are negative in direction.

Since sets function at an intermediate level, between motives and acts, the effects of the overall motivational pattern may override the effects of instructions. If someone who is nauseated is asked to act hungry, he may try and fail, or he may not even try. People are not always completely coöperative. If the problem is an artificial laboratory puzzle or decision, the thinker may readily adopt the requested set; but if the task has some vital significance, the thinker's persistent attitudes and values may prevent wholehearted compliance. If asked to judge the intellectual stature of political figures, for example, he may judge their politics instead. Many curious results have been obtained in the psychological laboratory because people are loath to adopt sets that conflict in any significant way with their ulterior motives.

A Previously Acquired Set May Be Reinstated. When a well-civilized person sits down to the dinner table, he is automatically prepared for eating or talking, or both, according to his habit. When he sits down in an automobile or airplane, a whole complex of specific reactions to specific signals is in readiness. It has been demonstrated, e.g., by Siipola (*1935*), that a set developed during solution of one problem in a series may be carried over, for better or worse, to the solution of subsequent problems of a similar nature. A man's reading habits or his occupational experiences may determine what he habitually looks for and what he is prepared to do when he meets any new problem or is required to make any decision. Many people deliberately aid the reinstatement of a set based on previous experience, as when the diner warns himself to watch for bones in the fish, or the chess player reminds himself to keep an eye on his opponent's bishop. Sets have a history of their own and presumably are subject to the same principles of learning, retention, and reinstatement as other activities of the same level of complexity and abstractness. The possibility of transfer and interference will have to be considered in connection with specific problem situations. We shall see, in later chapters, that sets acquired by previous experience play an important role in thought and judgment.

New Sets May Be Developed During the Performance of a Task. The development of new sets is of critical importance for the psychology of thought and judgment. Consider first the case in which a stimulus occurs for which someone is not prepared. If it is an intense stimulus, it will overcome the existing set and capture the subject's attention. Berlyne (*1950b*) has shown in a modification of the discrimination-reaction experiment that when confronted by two visual stimuli the subject without differential instructions will respond to the more intense, e.g., the larger, or the brighter.

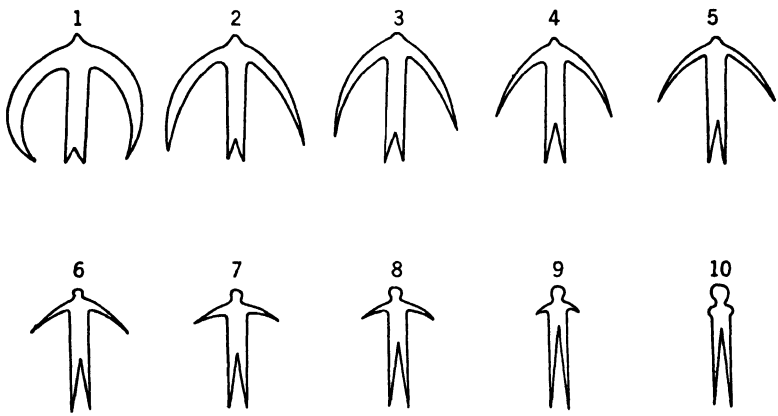


Figure 3.2. Drawings used by Hall (*1950*). See Table 3.1.

Usually, of course, people are somehow prepared for the task. But they modify the set slightly as they go along or, if their activities are unrewarding, shift set radically. Experiments on interference and distraction, to be considered later in this chapter, show that higher-order sets, encompassing originally disturbing stimuli, are often generated during performance of a continuous intellectual activity.

Wertheimer (*1923*) called attention to this way of generating a set in his experiments on perceptual organization. If a series of dot patterns is exposed, perception of the first few establishes a set which influences perception of the remainder. Perception of the patterns in the middle of the series will be reversed if the series is presented in reverse order.

Of several experiments illustrating the development of a set for naming drawings, one of the most interesting was recently reported by

Hall (*1950*). He used a series of ten outline drawings, those at either end being rather ambiguous while the middle ones could easily be perceived as a man (Figure 3.2). Eight different groups of ten subjects each were shown Drawings 2 to 9, one to each group, and asked to name them. The most important responses were "man" and similar names, like "doll" and "bird-man." The number of these manlike names for each group of ten subjects is shown in Table 3.1. Another group of ten subjects saw all drawings in forward order, and another group in reverse order. As the table shows, the manner of presentation made a significant difference in the naming. In the groups that saw all drawings, the naming of any one was influenced by the set developed while naming the preceding figures.

TABLE 3.1. Effect of Set Induced by Order of Presentation on the Naming of the Drawings Shown in Fig. 3.2.[a]

Order of	Drawings									
presentation	1	2	3	4	5	6	7	8	9	10
Shown singly		0	3	7	9	10	8	8	8	
Forward order	0	2	8	9	10	8	9	7	2	0
Reverse order	1	0	3	5	6	7	10	10	10	0

[a] The entries are the number of manlike names given to each drawing by three groups of ten subjects each. Data from Hall (*1950*).

Careful scrutiny of the table will disclose that the set has both positive and negative effects. Drawing 4, for example, was called manlike seven times when seen alone. When preceded by drawings 1 to 3 it was called manlike more often than this; when preceded by Drawings 10 to 5 it was called manlike less often. As we shall see later, judgment of lifted weights is influenced in the same way.

In the anagram problem the thinker has a very general set, established by the instructions, but the set may become more specific as he goes along. Rees and Israel (*1935*) made up a sequence of anagrams so that all could be solved by rearranging the letters in a certain order. The subjects were variable at first, but soon their responses showed that they had established a set for the order they had been using. A set for seeking a certain kind of word, e.g., a "nature" word, was established in the same way.

Postman and Crutchfield (*1952*) induced their subjects to adopt a set to produce food words as solutions to skeleton-word problems by starting them off on problems for which food solutions were obvious,

e.g., LUN–H, CHE–S–. The number of food solutions given to later problems of greater difficulty was clearly related to the number of such opportunities for getting set to think of food words. Different groups were given 0, 1, 2, 3, and 5 problems with obvious food solutions; average numbers of food solutions to the later problems were, respectively, 7.9, 8.7, 9.1, 9.6, and 10.3.

Immergluck (*1952*) established a set for judging pairs of geometrical designs for symmetry in the same way. The experimental group got 11 pairs of designs, the one on the right always being clearly more symmetrical. In the twelfth pair the positions were reversed, and 81 percent of these subjects followed their set automatically into error, as compared with a 13 percent error by control subjects who got the first 11 pairs in mixed positions.

Sets established by peculiarities of the task or sequence of tasks are especially important to problem solving. Whether the thinker can describe such sets or not, their operation may be inferred from his activities. It would not be quite correct to say that these sets are carried over from previous experience and reinstated as habits. They are developed in and for the task at hand. And it would not be correct to call them situational sets, because they depend on what the subject is doing as well as the situation he perceives. Maier (*1930*) has used the term "direction" for sets like these. Since these sets are similar to the others except in origin, they may conveniently be listed as *task-induced sets.*

In respect to subsequent activity, sets are preparatory. But at any moment sets are themselves activities of complex organisms. Hence they are influenced by many conditions in the individual and in his surroundings. The causation is multiple. Nevertheless, at any moment the organism is unified. He has one set. If he is set to turn right, he will not, as long as that set is in operation, turn left. This fact of unity is what makes the set important and its experimental investigation possible.

SETS IN OPERATION

The remainder of this chapter will be concerned with the effects of sets in operation, attempting to show how the sets established as described above influence human activities. The aim is to clarify the concept of set and to state some principles that can be applied in later chapters. Then we can consider some related topics that are important

in their own right, such as continuous intellectual work and mental fatigue. The remainder of this chapter could be called the applied psychology of set.

Effects on Perception and Attention. The set has a strong qualitative influence on perception, slanting the perceiver toward perception of some objects and some aspects of the environment rather than others. Anyone who is prepared, whether by verbal instructions, suggestion, hypnosis, a motive, or a habit, to hear a telephone ring is quite likely to hear a telephone ring even if it is the doorbell. If he is fearful of an intruder, he is likely to hear, see, or even smell an intruder. If he goes to bed in an anxious mood, he may dream about an intruder.

The afferent side of the set, the preparation for perceiving some things to the neglect of other things, often goes by the name attention (Paschal, *1941*). In some contexts this is what is meant by that protean word thought. To ask what someone is thinking about is to ask what he is attending to, what he is perceiving, feeling, imagining, anticipating, hoping for, or dreading. In addition to the influence of sensory stimulation from the environment, attention is influenced by the previously prepared set to attend.

It is relatively easy to direct anyone's attention this way or that by verbal instructions. If you ask someone, under ordinary conditions of social interaction, to look at the clouds, he probably will. The same effect can be produced indirectly, by what is called suggestion, if you hold out your hand, look upward, and scowl.

The concept of self-instruction is hard to pin down satisfactorily, but people will often claim that they deliberately pay attention to subtle aspects of the environment or to transitory events in spite of more compelling stimulation. And people will claim that they successfully tell themselves to cast evil or unpleasant thoughts out of their minds and replace them by cheerful thoughts. There is a vast popular literature on this subject, largely ignored by scientific psychology. Perhaps an extension of the concept of ego motivation will suffice for this phenomenon of "thought control," because the motives for the self-instruction are usually motives that the soul searcher is proud of. Attention directed in this way is often called voluntary attention.

Habits of attending to or thinking of things, however they are acquired, are well recognized, and are often called interests. Certainly a person's thoughts are influenced by his vocation, his avocations, and

his recent experiences. Odors may call to mind, because of past experience, a vision of a flower or fear of asphyxiation. For the same reason a familiar tune may turn one's thoughts toward dancing and fill another's mind with unpleasant memories.

This fleeting phantasmagoria of the mind does not interest psychologists today as much as in the past—though it still fascinates poets and novelists. Most modern textbooks of psychology consider thoughts in this sense under the heading of attention, listing briefly the determinants in the person, i.e., the set, and in the environment. Clinical psychologists and psychoanalysts have a practical interest in their patients' fantasies and attempt to work backward by a variety of subjective procedures from reported fantasies to unreported motives. Experimental psychology gets around this problem of subjectivity quite neatly by using set as an independent variable. The subject is asked to adopt a set for one observable response when he feels, imagines, or perceives one thing and another observable response for other things.

Külpe's (*1904*) classic experiment may be taken as an illustration of the influence on perception of experimental variation of the set established by instructions, though its methodological significance was perhaps not completely understood at the time. He used colored letters presented tachistoscopically and found that, when the subject was prepared to see color, he could report the colors of the letters but not much else. When he was set to see the letters or count them, he did so, but then was unable to report the colors. Many similar experiments with similar results have since been published (Moore, *1936;* Gibson, *1941;* Paschal, *1941;* Bruner and Postman, *1949*).

Postman and Bruner (*1949*) extended this experiment to demonstrate how the efficiency of perception is reduced when the set is a multiple one. They exposed pairs of words, such as yellow and fiddle, very briefly in a tachistoscope, recording total time necessary for correct recognition. In the single-set series the subjects were instructed to recognize a color word. In the multiple-set series they were told to perceive either a color word or a food word. Under the multiple-set instructions much more exposure time was necessary. The more complex set interfered with the efficiency of perception or, in the experimenters' words, reduced the "sharpness of tuning" of the set. A continuum of sharpness of tuning is conceivable, ranging from a set to perceive everything equally well (and therefore to perceive nothing

very well) to a set to perceive something maximally (and therefore to ignore everything else).

It follows from this discussion that conditions which facilitate performance on one perceptual task would interfere with performance on a competing perceptual task. Bahrick, Fitts, and Rankin (*1952*) demonstrated this principle by testing their subjects' perception of peripheral stimuli while they were tracking a moving target in central vision. Increasing the incentive for the tracking task reduced performance of the peripheral task.

Because of this influence of set on perception, it is standard practice in the psychological laboratory, when maximal perceptual efficiency is desired, to prepare the perceiver optimally. The precision of the results obtained in psychophysics demonstrates that verbal instructions will establish a sharply tuned set in coöperative, socially motivated human subjects. Analogous sets can be established in the lower animals by appropriate arrangements of apparatus and rewards.

Considering attention as the afferent side of the set ignores the response side. But though the response may not be specified by the instructions, response consequences can be observed. Preparation for perceiving includes an adjustment of the sense organs for optimal reception of the stimulus. For a visual stimulus the head is turned; the eyes come into focus and convergence. The head is turned for an auditory stimulus also, and some animals prick up their ears. For better perception of odors sniffing is common, and some animals wet their noses. There must be internal neural adjustments also, but of these little is known beyond the general organismic vigilance.

Conscious Thought as an Optimum. Many thinkers have assumed that the best thought is conscious thought, and to ignore consciousness is to amputate thought. Only when one is conscious of the critical feature of a problem is he able to solve it.

The reality behind this assumption can easily be stated in terms of set. When the thinker is set for the critical feature of a problem, he will attend to this feature, and this is what he will say he is conscious of. Only if he is set for the critical feature will he solve the problem with maximal efficiency. Consciousness of X, therefore, is not a determinant of good thinking about X. Both the consciousness (or attention) and the good thinking are consequences of a set for X.

Effects on Response. If someone sitting in a chair should see a light flash in front of him, his reaction, if any, would be quite un-

predictable. Different people would do different things. But in a reaction-time experiment the subject is set in advance, hence a high degree of predictability is obtained. The general instructions of the experimenter and his businesslike attitude, as well as the appearance of the laboratory, prepare the subject for something, alert him at least; then the experimenter's ready signal brings the preparation to a point. As a result of this preparation the subject will make a specific response to a light or buzzer, and will do so very quickly. He may, in fact, make the prepared response without the stimulus: the "false reaction." In consideration of the extreme complexity of the human organism, the tremendous number of stimuli in the environment, and the tremendous number of possible responses, the execution of one specific response a tenth of a second after one specific stimulus is a remarkable phenomenon, of great practical and theoretical significance.

If someone sitting idly in a chair hears the word "run," he might say or do anything, or nothing. One could not predict. But if he has been prepared by experimental instructions and by a little practice to name opposites, his response can be predicted with fair accuracy. If he casually lifts a small laboratory weight, he may make any kind of a response, or none. But if he has been asked to judge the weight "heavy" or "light," we may confidently predict that his response will be limited to these two words, and under some conditions we may predict the exact word. (See Figure 3.3.)

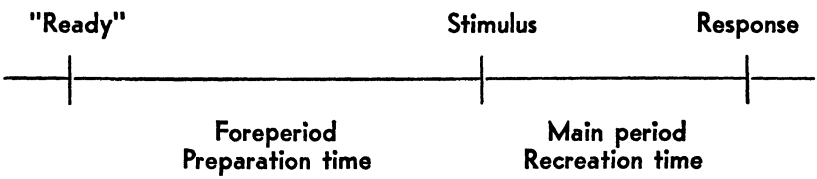


FIGURE 3.3. Time relations in the reaction-time experiment.

Preparing for reaction, or getting set, is an activity of the whole organism which takes a finite amount of time, is subject to experimental manipulation and control, and produces observable variations in the response. By systematically varying the foreperiod, i.e., the time between the ready signal and the stimulus, Woodrow (*1914, 1916*) found that the optimal preparation time for well-trained subjects was about two seconds. Foreperiods longer or shorter than this optimum produced longer reaction times on the average. When fore-

periods are irregular, i.e., when the subject cannot well prepare, reaction times are longer and one foreperiod is about as good as another. Following practice with foreperiods of a given length, e.g., 12 seconds, people will habitually adopt a set to expect the stimulus and make the response 12 seconds after the ready signal; hence, for the duration of the experiment, that foreperiod gives shorter reaction times than longer or shorter foreperiods (Mowrer, *1940*). Since these sets, established by verbal instructions, operate within the overall motivational pattern, when the usual social motivation is supplemented by electric shock, the set is strengthened and reaction time is shortened (Johanson, *1922*).

From the above data we might expect that if someone has to make a second reaction immediately after the first, a foreperiod of two seconds, or at least one second, would be necessary. Poulton (*1950*) has recently demonstrated, however, that the subject can prepare for the second while making the first, so that the time for the second is no longer than the usual reaction time. This can be done if the preparatory signal for the second reaction precedes the completion of the first by only four-tenths of a second. In Poulton's words: "The subject's anticipatory mental set or aim thus has a unifying effect, converting what might otherwise be a series of discrete responses to a succession of stimulus situations, into a single skilled performance." We shall return to this principle later in this chapter (pages 83–86).

Effects on Learning. The effect of an appropriate preparatory set in facilitating learning is well known (Dashiell, *1949;* Gibson, *1941*). An experiment by Moore (*1936*) shows how large this effect may be. He had a group of students read a paragraph, with instructions to remember as many ideas as possible. They remembered 51 on the average. Another group was instructed simply to read the paragraph. The average was 38. Another group, instructed to count the words, remembered an average of only 7 ideas.

Psychophysiology of Set. Several attempts have been made from time to time to record the neuromuscular events that occur when a laboratory subject, following instructions from the experimenter, gets ready to do something. Instructions emphasizing attention to forthcoming stimuli and readiness for intellectual tasks do have an influence on the alpha rhythm of the electroencephalogram, but the relationships are complicated and have not yet been clearly worked out (Lindsley, *1944;* Darrow, *1947*).

Records of muscular tension obtained by mechanical devices and records of electrical action potentials from the muscles have yielded more specific information on set (Davis, *1946;* Freeman, *1948*). When a laboratory subject is told to get ready to make a reaction, muscular tension in the reacting muscle usually increases (Freeman, *1931;* Davis, *1940*).

Davis compared the effect of the length of the foreperiod on reaction time, as demonstrated by Woodrow and Mowrer (see pages 76–77),

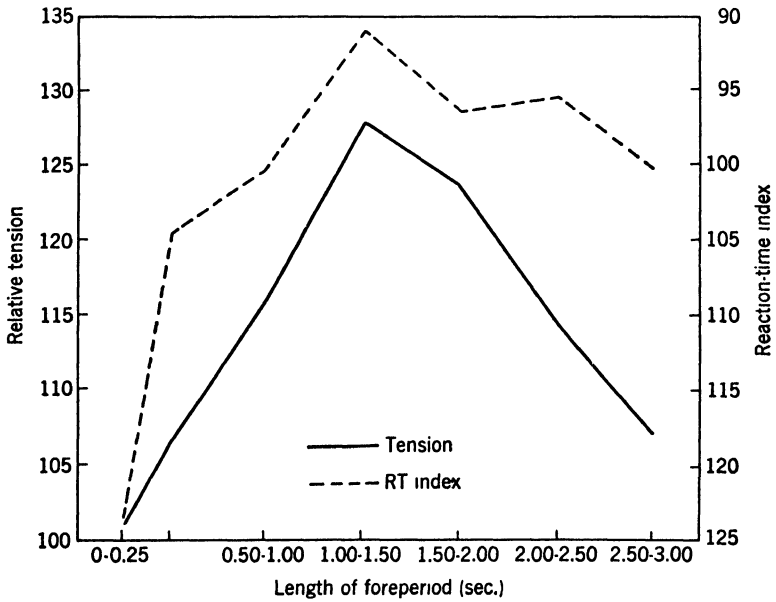


FIGURE 3.4. Showing reaction-time index and relative tension in relation to duration of foreperiod. From Davis (*1940*).

with changes in muscular tension during the foreperiod. As Figure 3.4 shows, the temporal course of action potentials in the principal muscle groups parallels the effect of verbally induced set in shortening reaction time.

A similar parallelism is noted when irregular foreperiods are used. Furthermore, both the facilitative effect of the set and changes in action potentials are influenced in the same way by practice.

When the subject adopts a motor set, emphasizing the speed of the reaction, the muscular tension is largely confined to the acting muscle

group; but when the subject adopts a sensory set, emphasizing perception of the stimulus, the preparatory muscular activity is spread out over more different muscle groups (Freeman, *1931*). During the foreperiod, if the experimenter asks the subject to change his set, the shift in muscular tensions may be clearly seen in the tracings (Figure 3.5). If muscular tension is externally induced at various intervals after the ready signal, an optimal interval is discovered, which is consistent with Woodrow's (see page 76) optimal foreperiod (Freeman and Kendall, *1940*). There is general agreement (Davis, *1946;* Freeman, *1948*) that the preparatory set is similar, in respect to distribution of tensions in different muscle groups, to the consummatory or goal response. The similarity is limited by the capacities of the organism (Katzell, *1948*).

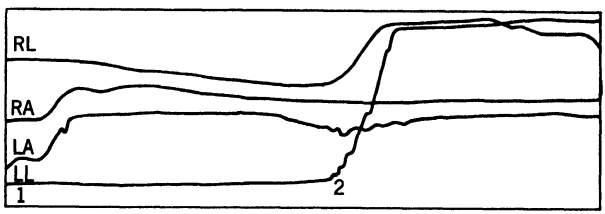


FIGURE 3.5. Effects of change of set on pattern of muscular tension. RA and LA refer to muscular tension in right and left arm muscles. RL and LL refer to muscular tension in right and left leg muscles. At 1 instructions were given to prepare for finger reaction. At 2 the subject was told to prepare for toe reaction instead. Traced from record by Freeman (*1931*).

From this brief survey of the evidence it is clear that the changes in muscular tension must be a part of the set. There are certain sense-organ adjustments also, corresponding to the perceptual or afferent side of the set, as pointed out above. And the evidence from electroencephalographic studies indicates the cortical involvement. All in all, the neuromuscular investigations of set agree with the definition of set and supplement the qualitative and quantitative effects of set on perception and response. When these various lines of research are pulled together in this way, *there is no longer any justification for the assertion that set is a vague, unmanageable concept.* We know how sets are

established. We know their effects on perception and performance. And we know a little about their physiology.

ATTENTION SPAN

The definition and most of the discussions of set have considered preparation of a response to one or two stimuli. Complex intellectual activities, however, require response to many different stimuli at once. Hence, the question arises in many contexts: How much can one get set for, or attend to? Any human being with normal sensory equipment can respond to many stimuli in the exciting world around him, many sights, sounds, and smells; but because of the unity of the organism, he cannot respond to all of them at once.

When attention is given an existential meaning, attention span means the number of objects simultaneously clear in consciousness. The question then can only be investigated introspectively, and the results have not been satisfactory. When attention is given a functional meaning of preparation for response to a number of items, objective experiment becomes feasible. The incoming or attention side of the preparation cannot be studied objectively unless a response is made, so the standard procedure is to ask the subject to make a very simple response to systematically varied numbers of stimulus objects.

It is admitted, without experiment, that anyone can attend to more things when he is prepared for them than when they come as a surprise, but experiment is necessary to determine how much one can attend to when he is optimally prepared. The question is one of upper limits, of spans. Tests of these spans under simple conditions have brought forth some empirical information that may be helpful in understanding more complex intellectual activities. The terminology is confusing, but the actual results are quite clear-cut.

In a typical test of the span of apprehension, several simple objects, e.g., dots, are presented briefly in a tachistoscope and the subject is asked to report how many he saw. The emphasis is on perception rather than response, and the score is often called the perception span as well as the attention span or apprehension span. In a typical test of immediate memory, e.g., the digit-span test, numbers are read to the subject and he is required to repeat exactly what he heard. Just as, in the perception span test, tachistoscopic presentation prevents repeated perception, the rapid presentation of the digits and the require-

ment of immediate reproduction prevent memorizing. Both tests, like the reaction-time test, call for a single burst of activity, the rapid execution of a prepared response. The reaction-time experiment asks how fast someone can run off a prepared response, while these two types of attention experiment ask how much someone can be prepared to respond to. As in the reaction-time experiments, the stimuli may be perceived by eye (span of apprehension), by ear (span of immediate memory), or by other sense organs.

Whatever the manner of presentation, span of attention is distinctly limited. Span of perception of dots for the average adult is about seven. Spans of immediate memory for disconnected digits and words are about the same. When the items are easily organized into groups, spans are larger if the number of items is counted, but not if the number of groups is counted. When words are organized into sentences, the number of words reproduced may be quite large. When the items are unfamiliar, the spans are reduced. Training increases spans for specific kinds of stimulus material.

There is much variation between subjects and between performances of one subject on different days, but the span is always limited to some small number. Attention to some stimuli interferes with attention to others. Hence span is reduced when activity is interpolated between perception and report (Pillsbury and Sylvester, *1940*). And span is greater for heterogeneous material, e.g., letters and digits, than for homogeneous material, e.g., either letters or digits (Young and Supa, *1941*).

As noted in the previous chapter, there is good evidence that perception of less than seven dots is a different process from perception of more than seven. It seems likely that immediate memory of less than seven digits is a different process from memorizing more than seven. Although a curve-fitting experiment analogous to those by which perception of visual numerousness was separated into subitizing and unitizing (page 39) might not be feasible, it is well known that time necessary for mastery of a list increases disproportionately when the list exceeds the span of immediate memory. When eight digits are presented, the subject may remember all of them; then when nine are presented, he may remember only six. The span of immediate memory seems to be a unit of attention. One can get set for and attend to all items within the span and hold them long enough to make a com-

pletely adequate response. Then, when the span is exceeded, additional and perhaps different operations are required. If this is true, correlations between scores for immediate memory and scores for memorizing lists beyond the memory span would not be high (see page 137).

A Hypothesis About Immediate Memory and Thinking. The span of immediate memory may be considered the number of disparate items that one can attend to or think about or get set to deal with in any way during any one moment. Or, in a negative way, no matter how many items one perceives, it is only those held in immediate memory that are responded to with maximal efficiency. It follows that during thought and judgment only those items that are in immediate memory can be attended to or thought about optimally.

Routine activity requires little attention. But if anyone has a problem, as that term is defined earlier in this chapter, the motivation for solving the problem will capture his set, directing attention toward the problem situation. Some problem situations are directly perceived by eye or ear, and some are abstract, involving words, ideas, concepts, and other symbolic material. In either case, whatever the items of the problem to be organized, manipulated, or otherwise dealt with, there is a limit to the number of separate items that can be thus grasped, retained, and manipulated, and that limit is the span of immediate memory. If remote memory is the storehouse of ideas, immediate memory is the workshop wherein ideas are processed.

This hypothesis has three important consequences. The first is that any intellectual activity is limited to consideration of only a small number of things at any one time, in most cases about five or six. If more than six items are involved in the problem, they will be grouped into a smaller number; or they will be considered sequentially, some relationships being neglected; or some of the items simply will not be considered. Actually, something like this does happen. When people are asked to decide a controversy on the basis of more than six separate arguments, the surplus arguments are usually ineffective. Good speakers, knowing the limitations of their audience, do not try to discuss more than five or six points at one time. If they have to make more than six points, they are careful to group them under a small number of headings. If the solution of a problem is contingent on manipulating more than six separate items simultaneously, the items will have to be organized in some way or the average person will not solve it.

Since the span of man's immediate memory has been approximately constant throughout recorded history, the historical divisions of any subject matter are usually some small number like six or seven: six continents of the world, seven wonders of the world, Seven Cities of Cibola, seven days in a week, seven ages of man, and so on. When there are more than seven, e.g., twelve months of the year, groupings may be seen, e.g., four seasons. If there is a natural division into two categories, such as day and night, subdivisions are introduced, such as forenoon, afternoon, evening, etc. When systematizers deliberately set up the divisions of a topic, the major divisions are usually about six. As any organization, whether industrial, governmental, or educational, expands and develops more than six or seven divisions, a tendency to combine several and reduce the number of major divisions soon appears. It is hard to believe that these facts are the results of historical accidents. It is more likely that man has divided up the facts of life in accordance with his ability to grasp them.

A second consequence of this hypothesis is that conditions which decrease span of attention will decrease the efficiency of intellectual operations. Fatigue, oxygen deprivation, alcoholic intoxication, and retroactive inhibition, for example, are known to reduce attention span. We would expect that these conditions would have a secondary disturbing effect on difficult problem solving.

A third consequence of the hypothesis relates to individual differences. If ideas can only be manipulated in immediate memory, those people with large memory spans will be better thinkers than those with small memory spans. A little positive evidence on this point will be brought forth at the end of this chapter.

SET FOR CONTINUOUS INTELLECTUAL WORK

Man is a unified mechanism, an organism, to be sure, but he is a complex harried organism, motivated in many ways toward many environmental patterns of changing stimulus value. Terms like set and attention come into use in psychology because the ulterior motivation of activity is not the same as its immediate control, just as, in a large industrial organization, the functions of the board of directors are not those of the plant manager. Even when the conditions initiating the activity, such as the motives and the instructions, remain constant, the activity varies. A thinker sits down to work on a problem with the

best will in the world, but his progress is usually crooked and discontinuous. He organizes his facts, getting off to a good start, then he wonders what is on at the movies tonight. He pulls himself back to the task at hand, gets a good hunch, and makes some progress, then he is distracted by dreams of reward for the solution. If the task is one in which progress can be recorded in small units, fluctuation will be objectively verified. We say that "his mind wanders" or "attention is blocked" or we speak of the variations of his set.

As ordinarily defined and as described above, set refers to the preparation for one burst of activity; but if the activity outlasts the reaction time, there must be something like a set that persists and integrates the flow of activity. Perhaps set is too static a term; steer has been suggested by some writers. Others speak of the control of attention. Attention in this usage has a functional rather than an existential meaning, but it is still too one-sided a term because we wish to refer to control of the response as well as what is responded to. Whatever the terminology, the data are objective measures of intellectual output over an extended period of time.

If the intellectual task is a complicated one, e.g., a problem to be solved, many complications are brought in. Hence simple or well-practiced tasks are employed: canceling *o's,* naming colors, performing familiar arithmetical operations, reading easy material, and the like. We are not here concerned with the determinants of any particular response to any particular stimulus, but rather with the control of the overall pattern of activity, the total output. Assuming that the subject is motivated for high output, the question becomes one of efficiency. We would like to know why his output fluctuates from time to time and why he stops when he does.

In the reaction-time experiment we may ask someone to make one respond if a red light flashes and another if a green light flashes—one burst of activity and he is done. Suppose now that we ask him to say "red" when he sees a red square and "green" when he sees a green square. And suppose further that we print red and green squares on paper in irregular sequence and instruct the subject to name the color of the first one, then the color of the next, and so on, as fast as he can. The subject will not have to be instructed for each response individually. He moves right along, responding to each

stimulus according to instructions, and makes quite rapid progress. If we ask him to read letters aloud from a few lines of this book, we may find that he can respond adequately to 150 letters a minute, averaging one response every 0.4 second. The separate acts are varied and irregular, but are integrated by the continuing set in such a way that the activity as a whole proceeds quite smoothly.

When someone names colors, reads letters or words, or cancels *o's,* his activity must be controlled by a discriminative set, as in the discriminative reaction-time experiment, since he responds differentially to the different stimuli. What he responds to after he gets well into his stride is a stimulus pattern which includes the preceding completed response perceived visually or kinesthetically. In a well-practiced activity like reading aloud, the vocal responses are geared to each other as well as to the visual stimuli. The eye-voice span may be five or six words. We may say either that the response mechanism lags six words behind the perceiving mechanism or that the reader is prepared for the next six words ahead.

Experiments on briefly exposed stimuli cited earlier show that one perceives what he is prepared for, neglecting the remainder. The same principles applies to continuous activity. When one is driving a car, he watches for turns, pedestrians crossing the road, stop signs, and traffic policemen, neglecting shadows on the road, dirt on the windshield, and pedestrians on the sidewalk. In reading one looks for meanings, ignoring defects in the paper, imperfect printing, and perhaps even typographical errors. But if the task is to cross out all the *t's* on a page, the meaning will probably be ignored.

INTERNAL INTERFERENCE

Many years ago Culler (*1912*) demonstrated the effects of interference on speed of continuous sorting of cards. A control group of eight subjects sorted the cards into pigeonholes eight times, picking up speed as they went along. A group of fourteen subjects sorted four times, picking up speed in the same way, then were shifted to another sorting system for four trials. Returning to the first system, they were a little faster than when they started, slower than they were on the fourth trial with the first system, and much slower than the control group on their fifth trial. Table 3.2, extracted from Culler's report, illustrates some of the pertinent results.

TABLE 3.2. Effects of Interference in Card Sorting[a]

Control group	119	100	93	93		83	85	83	79
Experimental group	116	101	96	91	Four trials with another system	112	96	90	85

[a] This shows the average time per trial for a control group of 8 that sorted eight times without interference and an experimental group of 14 that sorted before and after interference. The interference consisted of sorting by another system. From Culler (*1912*).

Both of these groups and other groups who shifted in other ways improved considerably over a six-day period, but the control group remained the fastest. The difference, of course, is not in the separate acts, which were practically identical, but in shifting the set. The experimental subjects were set to put each card into its pigeonhole, then had to adopt another set to put each card in some other pigeonhole. In general, practicing one set interferes with the efficiency of an incompatible set, but the whole complex activity improves with prolonged practice.

When the two tasks that are alternated are distinctly dissimilar, little or no interference occurs. Jersild (*1927*) had his subjects alternate between simple arithmetic operations and controlled association, and found no drop in speed whatever as compared to continuous performance of either task alone. When they alternated between two similar arithmetic operations or between two similar associations, the expected difference in speed appeared.

This internal interference in continuous intellectual work is similar to the internal inhibition in immediate memory reported by Young and Supa (see page 81). But in continuous intellectual work the interference is between little sets for specific stimuli and responses which may later be integrated, as noted above, into a compound or higher-order set for continued activity. And this higher-order set to continue according to instructions may or may not be disturbed, depending upon conditions to be discussed below.

DISTRACTION

By distraction we mean that someone set for continuous intellectual work is stimulated by events external to the task, with the result that he works less efficiently. Events of particular interest to the subject and stimuli similar to the subject's task are especially distracting.

The discussion of internal interference noted that stimuli which in-

itially disrupt the set may be later dealt with by the development of a more inclusive task-induced set, and the same higher-order integration may encompass initially distracting stimuli. Whatever the forces of interference or distraction may be, they are effective only if applied to the critical point at the critical time, i.e., to the response that is prepared when it is prepared. Reconstructing the set, so that the stimulus pattern one is prepared for includes the previously distracting stimuli, cuts out the effect of the distraction.

Thus, if someone expects that music will reduce his performance, it is likely to do so, and the reverse expectation may reverse the results (Baker, *1937*). This principle cannot be carried too far, however. Most college students would say, if asked, that noisy working conditions will reduce scores on an intelligence test, yet in Hovey's (*1928*) experiment the reduction was hardly noticeable. And in an experiment on the effect of music on reading difficult material (Henderson, Crews, and Barlow, *1945*), those who said that they like to study with the radio on were not immune to its distracting effects.

THE PSYCHOPHYSIOLOGY OF CONTINUOUS INTELLECTUAL WORK

From the short-time point of view, as in the reaction-time experiment, muscular tension and similar physiological measures are signs or indicators of readiness for a burst of activity. The better the preparation, the better the response. Over a longer span of time muscular tension may be taken as an index of effort or physiological cost of work. It was once assumed in discussions of interference and distraction that the subject exerts constant effort, hence that an increase in output is an increase in efficiency. This assumption was called into question by Morgan's (*1916*) investigation, which showed that well-motivated people, when distracted, are likely to maintain or even increase output by exerting more effort. The increased effort is indicated by greater muscular activity, changes in breathing, and speech movements. More oxygen is consumed at first, but later—presumably when a higher-order set is developed—oxygen consumption decreases (Harmon, *1933*). As muscular tension, at first distributed throughout many muscles, becomes more focalized, the metabolic cost of work decreases (Freeman, *1939*).

Continuous intellectual work, as in arithmetic and substitution tasks, is improved somewhat by ingestion of Benzedrine sulfate (Kleemeier

and Kleemeier, *1947*). The improvement is most noticeable on tasks that require flexibility in shifting set.

In the Minnesota experiments on severe vitamin-B deficiency, the loss in intellectual performance was surprisingly small. The only significant loss was in the Flags and Multiplication Tests, both of which, when well practiced, are speed tests requiring continuous attention (Guetzkow and Brozek, *1946, 1947*).

From these and other experiments we may be sure that the energy required for intellectual work is very small. When intellectual output declines, the difficulty is one of mobilization of energy, of alertness. From the engineering point of view we might say that the problem is one of signaling, or control, rather than one of the generation and conservation of power.

All this raises the important practical question whether muscular tension could be used as a measure of alertness and a warning when alertness is dangerously low. Travis and Kennedy (*1947*) devised electronic apparatus that could accumulate electrical potentials from electrodes placed on the forehead just above the eyebrows, and gave their subjects tasks that required continuous attention. As expected, errors were more frequent when muscular tension levels were low.

Next, they (Kennedy and Travis, *1948*) adjusted their apparatus so that when the muscular action potentials fell below a certain level, a light would flash to which the subject had to react quickly. If the stimulus came when the subject's tension level was high, i.e., when he was alert, average reaction time was short; but if the stimulus came when tension level was low, average reaction time was much longer and failure to react was much more frequent.

INTELLECTUAL SATIATION AND FATIGUE

The longer anyone persists in any kind of work, the more vulnerable he becomes to interferences of all sorts. If it is easy, repetitive work, he will often complain of fatigue or boredom after an hour or so. His output will fall off in quantity and quality. He may relax his vigilance, judging his work by a lower standard (Mackworth, *1948*). And he may quit altogether. There is no single explanation for this phenomenon that applies to all cases, but a number of contributing factors can be briefly outlined.

To understand why anyone stops work, we should ask first why he started. If the motivation is an extrinsic biological or social one, he

may slow down or quit simply because his motivation decreases, or because the activity does not seem to be biologically or socially rewarding. Under many conditions, inside the laboratory and outside, if there is greater reward for high output than for low output, it is often remote and intangible. A continuous intellectual task like mirror writing or making gates may offer a challenge to the ego for a few minutes, but as soon as one learns how well or how poorly he can do it, he is satisfied. A complicated puzzle may be intrinsically attractive until it is solved, but the intrinsic appeal of easy, continuous intellectual activities is short-lived. Anyone is likely to slow down and stop when there is no good incentive to keep going. When working anagram problems, one is likely to continue in a line of activity if he is confident of success and if no alternatives are apparent (Robinson, *1940*).

In addition to lack of incentive to continue, new resistances to continuance may develop and overcome the motivation. Whatever the cause, attention to the task at hand becomes more difficult. The set, initially unified, is fragmented. This difficulty is more prominent in repetitive work than in work that is varied. Poffenberger (*1927*), for example, found that his subjects could continue working on intelligence tests, one form after another, for five and a half hours, improving about 20 percent from beginning to end. But when the task was simple repetitive addition, all ten of his subjects became steadily worse, and all gave out completely in less than five hours.

Aside from this difference between varied and repetitive work, generalization is difficult. What is monotonous to some people is fun to others. A task that is pleasantly varied for some produces too much interference for others. Someone may be satiated on simple repetitive acts but perform them readily when they are included as parts of a larger total activity (Karsten, *1928*). Reading in some respects is a simple, repetitive act which requires good control of attention and permits little large-muscle activity, yet people do it for hours at a time without complaint if conditions are otherwise favorable. A wide variety of laboratory and industrial experiments furnish specific illustrations (Poffenberger, *1942;* Bartley and Chute, *1947*), but the basic dynamics are not well understood.

One thing is clear: strong motivation can overcome almost any resistance to intellectual work. Two remarkable experiments demonstrate this point and add some interesting details.

In 1909 Tsuru Arai (*1912*), a graduate student working with E. L. Thorndike, attacked this problem by mutiplying four-place numbers, such as 7362 x 3594, mentally for 12 hours at a stretch. To control the practice effect, she practiced for several days in advance until she believed she had reached her limit of improvement. Then she started at 11 A.M. and worked until 11 P.M. steadily, stopping only for a few seconds after each problem to write down the answer and the time. She did this for four consecutive days, holding herself to 67 problems a day. Her pace slowed down somewhat during the course of each 12-hour day, and the number of errors increased from day to day. She got 21 right out of 67 the first day, 14 the second, 14 again the third, and only 10 the last day. According to one estimate, each problem requires a hundred intellectual operations; according to any estimate the intellectual effort is tremendous, even if errors are made. A few years later Painter (*1916*) tried the same feat, but gave out suddenly after four hours.

The experiment was repeated in 1930 by three women, Huxtable, White, and McCartor (*1949*), all graduate students at the University of Southern California. Like Arai their pace slowed down slightly during each 12-hour day and their errors increased from day to day. From beginning to end each one speeded up, however, even though they had all practiced thoroughly in advance. Average number of problems done the first day was 74, the second day 92, the third day 104, and the last day 102.

Along with this decrement in performance went minor physiological disturbances. Metabolic rate increased, actone was found in the urine, and all subjects lost weight. These changes were small, and could be due largely to omission of the midday meal. On the subjective side intense dislike and boredom were reported, as well as headache, confusion, distractibility, and irritability. One subject said that she "would not repeat those four days for $10,000."

Evidently, then, prolonged intellectual work near the upper limit of capacity can be carried on continuously for long periods. The physiological fatigue effects are negligible as compared with fatigue effects from muscular work. Perhaps the subjective effects were greater, though comparison is difficult. The chief contribution of these two studies is the demonstration of what can be done when motivation is high.

INDIVIDUAL DIFFERENCES

This chapter has demonstrated that thought and judgment depend upon many dynamic factors. And people may differ one from another in respect to any of these factors. Consequently, if our treatment of thought and judgment aims at completeness, we must consider individual differences in respect to all factors mentioned in this chapter.

MOTIVATION. Differences between people in primary motivation are not easily measured. Adults at least are motivated toward a goal by many positive and negative, conscious and unconscious, factors of direct and circuitous origin. Hence it turns out that it is easier to describe and differentiate people in respect to their goals than to determine why, in any ultimate sense, they strive for these goals. We have considerable information about individual differences between people in respect to goals, wants, likes and dislikes, interests, and values, obtained chiefly by asking people to check certain phrases collected in printed inventories of possibilities. Some of the items in interest inventories seem to get at intrinsic motivation of an intellectual sort, asking people, e.g., if they like to work puzzles; and high scores on the theoretical value of the Allport-Vernon Study of Values appear to indicate intrinsic motivation toward abstract thought. Vocational interests and educational interests show moderate correlations with achievements in these fields (Fryer, *1931*).

SET. When the task is a short one, a brief response to a briefly exposed stimulus, certain individual differences in the set, however established, are significant determiners of the results. In many of the laboratory experiments on set, different results are obtained by instructing two groups of subjects differently or by subjecting them to different previous experiences. Usually these sets are born and die in the laboratory; but when suitable stimuli are exposed briefly, the effects of stable individual differences can be demonstrated.

The reaction-time experiment gives reliable measures of individual differences in speed of simple reaction if enough reactions are timed. Individual differences in receptor function are rather specific, as are individual differences in response function; it is reasonable to suppose that correlations between different reaction times depend chiefly upon a common central readiness or preparation for activity. One would expect, then, that the intercorrelations between various complex reaction times which depend largely upon efficiency of preparation would be higher than the intercorrelations between various simple reaction

times. Such evidence as there is, which does not clearly answer this question, has been summarized by Woodworth (*1938*).

When the reaction-time test is given in such a way as to test ability to maintain set, schizophrenics show a definite defect (Huston, Shakow, and Riggs, *1937;* Rodnick and Shakow, *1940*).

Attention Span. Span of apprehension has not been widely used as a test of individual differences, but tests of immediate memory span have been found to be quite useful. Tests of digit span and immediate memory for sentences correlate well with the general factor of intelligence tests and are used in several individual test batteries.

Immediate memory spans are longer for some kinds of material than others, but scores on different tests correlate highly. Centroid factor analysis indicates the operation of a large general factor (Brener, *1940*). Ability to perceive spatial patterns and reproduce them immediately appears to be a general ability, regardless of mode of presentation (Compton and Young, *1933*) or mode of reproduction (Thomas and Young, *1942*). Evidently, then, the amount that one can get set for in a short time interval depends upon a general factor of some significance.

If our hypothesis about immediate memory span and thinking is correct, people with small spans would have difficulty in solving problems or making judgments that require attending to many objects, facts, or inferences at once. As compared with those who have large spans, they would be more liable to errors of omission, and would prefer conceptual systems of two or three divisions. One would expect furthermore that backward memory span would be a more useful test than forward span because it comes closer to the operations involved in intellectual manipulation. If we accept score on a test of general intelligence as a measure of ability to think, some positive evidence is available, as noted above.

Continuous Attention. Many psychological tests require rapid, continuous, easy intellectual work. As such they may be assumed to depend in part on individual differences in ability to attend. People who cannot concentrate for ten minutes or are easily distracted will get low scores. Tests of name and number checking, digit-symbol substitution, easy reading, simple arithmetic, following directions, and the like require continuous performance of easy tasks in irregular order. From inspection one would assume that such tests, when well practiced, depend in large part on control of set, development of com-

pound sets, and resistance to blocking and distraction of attention. Tests of this kind have often been called tests of attention. They have often been used in the selection of personnel and occasionally in research on fluctuations of efficiency. Scores on almost any psychological test in which speed on easy tasks is important would appear to test control of attention or set to some extent.

Going beyond inspection, a few attempts have been made to analyze tests in this area by modern correlation methods. One of the most thorough attempts has been reported by Wittenborn (*1943*). Building on previous work, he laid down four requirements for tests of attention:

1. The performance should not depend too much upon intellectual level.
2. The tasks should depend to as small a degree as possible upon content and knowledge.
3. The tests should correlate as little as possible with factors heretofore identified.
4. The scores on the tests should depend to a large degree upon a continuous, sustained application of mental effort. The tests should be so constructed that a layman might say they required a high degree of continuous "concentration."

On these principles he constructed six new tests of attention, to be presented by means of a phonograph record. These and other possible attention tests and tests of known factorial content, 20 in all, were administered to 175 men in an Air Force Technical School. Seven centroid factors were extracted and after rotation one was rather easily identified as an attention factor. It included a test of following directions, and the six new attention tests. That the best test of this factor requires strict concentration will be apparent from the instructions. The subject hears three digits. He is asked to put a plus sign in the appropriate space of his answer sheet (1) if the first digit is the largest and the second is the smallest, or (2) if the third is the largest and the first is the smallest. Otherwise he is to leave the space blank. These three-digit numbers are read rapidly.

Other good tests of Wittenborn's attention factor use consonants and vowels, numbers and letters, singly and in combination. Apparently the attention factor is not one of content or manner of presentation, but rather of "mental effort."

Several new tests, constructed by Army Air Force psychologists

(Guilford, *1947*) and called Integration Tests, are particularly interesting at this point. The rationale of these tests is described in the following paragraph:

> Analysis of the performance required of the pilot trainee in primary school, or of the navigator or bombardier in advanced school, suggests that the successful air-crew member must maintain sets to respond to a large number of conditions, cues, and reference points. Often these conditions must be observed simultaneously or at least within a brief period of time. Moreover, they must often be noted while some other activity is being carried out, thus making it necessary to divide attention without disrupting the pattern of action in progress. Some cues, when they occur, call for immediate action; others call for delayed responses with which there must be no interference by intervening activities. In order to respond appropriately, the various conditions that influence action at a given time must be observed, remembered, and integrated. . . .

Several integration tests were constructed on these principles and administered, along with familiar tests, 32 in all, to 266 classified pilots. This is a selected sample, of course, because they had all passed a battery of classification tests. Factor analysis by the centroid method identified three integration factors. The nature of one integration factor can be surmised from a description of the best test of this factor, Signal Interpretation:

> In each problem there is a diagrammatic representation of 10 airplane carriers in a row. The examinee must determine from which of these carriers planes will take off. This can be ascertained by interpreting certain signals such as the number of flags on the mast of each ship, the direction the ship is heading, and whether there are more or fewer flags than on the previous take-off ship.
>
> The examinee must compare quickly the number of flags on each ship with the number on the previous take-off ship. He must decide from the relationship between these two the direction of the next take-off ship.

It is definitely a speed test since the average pilot gets about 100 items right in the allotted 7 minutes. This factor "seems to represent the effective persistence of a complicated mental set which operates in rapid, complex, clerical-type work."

The second integration factor had its highest loading on a test of following directions, which appeared also in the Wittenborn studies mentioned above as a test of attention. It "seems to represent an adaptability of mental set: the trait of being able to modify sets on short

notice. One might be tempted to call it flexibility of set (absence of perseveration), except for the fact that flexibility (or perseveration) tests have notoriously failed to intercorrelate to any substantial degree."

The tests of the third integration factor required in general taking a number of considerations into account in making quick decisions. Neglecting any of the details of the instructions or the test material would lead almost inevitably to an incorrect response. "The third factor may represent some kind of span or scope of apprehension or attention: the ability to keep all elements in a set operating effectively."

Evidently the AAF integration tests have much in common with other tests of individual differences in control of attention or set. None of these tests was a pure test of any integration factor; almost all were correlated with other integration factors even in this selected sample. They correlated also with the general reasoning factor, to be described in Chapter 11, with the verbal factor, and to some extent with the space and number factors. These tests contributed some independent validity to the pilot criterion.

All in all, the evidence at hand indicates that a common factor can be identified in tests that require continuous attention, maintaining a set, or following irregular directions. As such it is of some importance in continuous intellectual work. But it is a narrow factor that probably has little significance for the solution of single untimed problems.

PERSISTENCE. As stated above, the longer anyone works on a simple repetitive task, the more likely he is to slow down and quit. Whether the task is easy or hard, some people will quit before others. The question of a factor of persistence has therefore been raised by several investigators, and some ingenious measures of such a factor have been devised.

Ryans (*1938a*), for example, used a number of objective tests of persistence, along with ratings and other data on 40 college students. The intercorrelations were factored by the centroid method, but the factor loadings were not rotated. However, the first centroid factor, defined quite well by the ratings of persistence and by a log of study time kept for a week, included some interesting measures of persistence. One score was the time spent attempting to decipher a very difficult code. Another was the time spent on an anagram task before quitting, and another was the number of words constructed. Another, called continuous mental work, required working addition problems for 40 minutes, the score being the number correct in the last 8

minutes as a fraction of the number correct in the first 8 minutes. The results suggest, in general, that success on continuous intellectual tasks is influenced by a persistence factor that is only slightly related to general intelligence. There is some evidence that persistence tests are useful in predicting college grades (Ryans, *1938b;* Thornton, *1940;* French, *1948*).

Thornton's (*1939*) factor analysis of persistence tests was carried out with 22 variables from 189 college students. One centroid factor identified after rotation was called "willingness to withstand discomfort in order to achieve a goal." A second factor, described as patience or "willingness to spend time at a task," seems to be the same as Ryans' principal factor. It included time spent on an anagram problem and the number of words constructed. An interesting reading task was also included in which perception of the words became increasingly difficult as the reader progressed. Time spent on the test and amount read correlated with the patience or plodding factor.

The time spent by students on their final examinations is an easily obtained datum that could be taken as an indication of this persistence factor. It has been demonstrated in three psychology classes (Briggs and Johnson, *1942*) that the first third of the students to hand in their papers get lower scores than would be expected from their intelligence, while the last third get higher scores than expected.

A factor analysis by Rethlingshafer (*1942*) was based on 29 variables involving persistence and other measures of continuance of activities. Although the analysis was based on the scores of only 38 subjects, it does agree with previous work in identifying the factor of keeping at a task once started. The factor of willingness to endure discomfort also appeared, as in Thornton's analysis, chiefly in duration of grip on the finger dynamometer.

Other analyses of tests in this area have been catalogued by Cattell (*1946*). The picture is not yet clear, but the evidence points to a persistence factor that determines the amount of time people will spend on various tasks, and thus the amount they will accomplish, when the instructions about stopping are indefinite. Persistence seems to be related to the will or character factor. Certainly the persistence factor makes an important contribution to achievement in any intellectual activity that lasts very long. According to all the evidence, it is independent of the content of the tasks and independent of general intelligence. It seems to be different also from such related factors as

perseveration, rigidity, and willingness to endure physical strain or discomfort, but these latter factors probably play a part in achievement on some intellectual performances.

REFERENCES

Allport, G. W., 1937. *Personality,* New York, Henry Holt and Co.

Arai, T., 1912. *Mental fatigue,* Teachers College Contributions to Education, No. 54, New York, Columbia University.

Bahrick, H. P., Fitts, P. M., and Rankin, R. E., 1952. Effect of incentives upon reactions to peripheral stimuli, *J. Exper. Psychol., 44*:400–406.

Baker, H. H., 1937. Pre-experimental set in distraction experiments, *J. Gen. Psychol., 16*:471–486.

Bartley, S. H., and Chute, E., 1947. *Fatigue and impairment in man,* New York, McGraw-Hill Book Co.

Berlyne, D. E., 1950a. Novelty and curiosity as determinants of exploratory behavior, *Brit. J. Psychol.* (General Section), *41*:68–79.

Berlyne, D. E., 1950b. Stimulus intensity and attention in relation to learning theory, *Quart. Jour. Exper. Psychol., 2*:71–75.

Brener, R., 1940. An experimental investigation of memory span, *J. Exper. Psychol., 26*:467–482.

Briggs, A., and Johnson, D. M., 1942. A note on the relation between persistence and achievement on the final examination, *J. Educ. Psychol., 33*:623–627.

Bruner, J. S., and Postman, L., 1949. Perception, cognition, and behavior, *J. Personality, 18*:14–31.

Cattell, R. B., 1946. *Description and measurement of personality,* Yonkers, World Book Co.

Compton, R. K., and Young, P. T., 1933. A study of organic set: immediate reproduction of spatial patterns presented by successive points to different senses, *J. Exper. Psychol., 16*:755–797.

Culler, A. J., 1912. *Interference and adaptability,* Arch. Psychol., No. 24.

Darrow, C. W., 1947. Psychological and psychophysiological significance of the electroencephalogram, *Psychol. Rev., 54*:157–168.

Dashiell, J. F., 1940. A neglected fourth dimension to psychological research, *Psychol. Rev., 47*:289–305.

Davis, R. C., 1940. *Set and muscular tension,* Bloomington, Indiana University Science Studies, No. 10.

Davis, R. C., 1946. The psychophysiology of set, in *Twentieth century psychology,* ed. P. L. Harriman, New York, Philosophical Library.

Davis, R. T., Settlage, P. H., and Harlow, H. F., 1950. Performance of normal and brain-operated monkeys on mechanical puzzles with and without food incentive, *J. Genet. Psychol., 77*:305–311.

Freeman, G. L., 1931. The spread of neuro-muscular activity during mental work, *J. Gen. Psychol., 5*:479–494.

Freeman, G. L., 1939. Changes in tension-pattern and total energy expenditure during adaptation to "distracting" stimuli, *Am. J. Psychol., 52*:354–360.

Freeman, G. L., 1948. *Physiological psychology,* New York, Van Nostrand Co.

Freeman, G. L., and Kendall, W. E., 1940. The effect upon reaction time of muscular tension induced at various preparatory intervals, *J. Exper. Psychol., 27*:136–148.

French, J. W., 1948. The validity of a persistence test, *Psychometrika, 13*:271–277.

Fryer, D., 1931. *Measurement of interests,* New York, Henry Holt & Co.

Gibby, R. G., 1951. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables, *J. Projective Techniques, 15*:3–26.

Gibson, J. J., 1941. A critical review of the concept of set in contemporary experimental psychology, *Psychol. Bull., 38*:781–817.

Guetzkow, H., and Brozek, J., 1946. Intellectual functions with restricted intakes of B-complex vitamins, *Am. J. Psychol., 59*:358–381.

Guetzkow, H., and Brozek, J., 1947. Intellective tests for longitudinal experiments on adults, *Am. J. Psychol., 60*:350–366.

Guilford, J. P., 1947. Printed classification tests. *Army Air Forces Aviation Psychology Program Research Reports,* Report No. 5. Washington, D.C., Government Printing Office.

Hall, K. R. L., 1950. Perceiving and naming a series of figures, *Quart. J. Exper. Psychol., 2*:153–162.

Harlow, H. F., 1950. Learning a satiation of response in intrinsically motivated complex puzzle performance by monkeys, *J. Comp. Physiol. Psychol., 43*:289–294.

Harlow, H. F., Harlow, M. K., and Meyer, D. R., 1950. Learning motivated by a manipulation drive, *J. Exper. Psychol., 40*:228–234.

Harmon, F. L., 1933. *The effects of noise upon certain psychological and physiological processes,* Arch. Psychol., No. 147.

Henderson, M. T., Crews, A., and Barlow, J., 1945. A study of the effect of music distraction on reading efficiency, *J. Appl. Psychol., 29*:313–317.

Hovey, H. B., 1928. Effects of general distraction on the higher thought processes, *Am. J. Psychol., 40*:585–591.

Huston, P. E., Shakow, D., and Riggs, L. A., 1937. Studies of motor function in schizophrenia: II. Reaction time, *J. Gen. Psychol., 16*:39–82.

Huxtable, Z. L., White, M. H., and McCartor, M. A., 1949. *A re-performance and re-interpretation of the Arai experiment in mental fatigue with three subjects,* Psychol. Monogr., 59, No. 275.

Immergluck, L., 1952. The role of set in perceptual judgment, *J. Psychol., 34:*181–189.

Jersild, A. T., 1927. *Mental set and shift,* Arch. Psychol., No. 89.

Johanson, A. M., 1922. *The influence of incentive and punishment on reaction time,* Arch. Psychol., No. 54.

Karsten, A., 1928. Psychische Sättigung, *Psychol. Forsch., 10:*142–254.

Katzell, R. A., 1948. Relations between the activity of muscles during preparatory set and subsequent overt performance, *J. Psychol., 26:*407–436.

Kennedy, J. L., and Travis, R. C., 1948. Prediction and control of alertness. II. Continuous tracking, *J. Comp. Physiol. Psychol., 41:*203–210.

Kleemeier, L. B., and Kleemeier, R. W., 1947. Effects of benzedrine sulfate (amphetamine) on psychomotor performance, *Am. J. Psychol., 60:*89–100.

Külpe, O., 1904. Versuche über Abstraktion, *Ber. I. Kongr. Exper. Psychol.,* 56–68.

Levine, R., Chein, I., and Murphy, G., 1942. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report, *J. Psychol., 13:*283–293.

Lindsley, D. B., 1944. Electroencephalography, in *Personality and the behavior disorders,* ed. J. McV. Hunt, New York, Ronald Press.

McClelland, D. C., and Atkinson, J. W., 1948. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception, *J. Psychol., 25:*205–222.

McClelland, D. C., Atkinson, J. W., and Clark, R. A., 1949. The projective expression of needs: III. The effect of ego-involvement, success, and failure on perception, *J. Psychol., 27:*311–330.

McGinnies, E., 1949. Emotionality and perceptual defense, *Psychol. Rev., 56:*244–251.

Mackworth, N. H., 1948. The breakdown of vigilance during prolonged visual search, *Quart. J. Exper. Psychol., 1:*6–21.

Maier, N. R. F., 1930. Reasoning in humans: I. On direction, *J. Comp. Psychol., 10:*115–143.

Meltzoff, J., 1951. The effect of mental set and item structure upon response to a projective test, *J. Abnorm. & Social Psychol., 46:*177–189.

Moore, J. H., 1936. The role of determining tendencies in learning, *Am. J. Psychol., 48:*559–571.

Morgan, J. J. B., 1916. *The overcoming of distraction and other resistances,* Arch. Psychol., No. 35.

Mowrer, O. H., 1940. *Preparatory set (expectancy)—some methods of measurement,* Psychol. Monogr., 52, No. 233.

Painter, W. S., 1916. Efficiency in mental multiplication under extreme fatigue. *J. Educ. Psychol., 7:*25–30.

Paschal, F. C., 1941. The trend in theories of attention, *Psychol. Rev., 48:*383–403.

Pillsbury, W. B., and Sylvester, A., 1940. Retroactive and proactive inhibition in immediate memory, *J. Exper. Psychol., 27*:532–545.

Poffenberger, A. T., 1927. The effects of continuous mental work, *Am. J. Psychol., 39*:283–296.

Poffenberger, A. T., 1942. *Principles of applied psychology,* New York, Appleton-Century-Crofts.

Postman, L., and Bruner, J. S., 1949. Multiplicity of set as a determinant of perceptual behavior, *J. Exper. Psychol., 39*:369–377.

Postman, L., and Crutchfield, R. S., 1952. The interaction of need, set, and stimulus-structure in a cognitive task, *Am. J. Psychol., 65*:196–217.

Poulton, E. C., 1950. Perceptual anticipation and reaction time, *Quart. J. Exper. Psychol., 2*:99–112.

Rees, H. J., and Israel, H. E., 1935. *An investigation of the establishment and operation of mental sets,* Psychol. Monogr., 46, No. 210.

Rethlingshafer, D., 1942. Relationship of tests of persistence to other measures of continuance of activities, *J. Abnorm. & Social Psychol., 37*:71–82.

Robinson, E. E., 1940. An experimental investigation of two factors which produce stereotyped behavior in problem situations, *J. Exper. Psychol., 27*:394–410.

Rodnick, E. H., and Shakow, D., 1940. Set in the schizophrenic as measured by a composite reaction time index, *Am. J. Psychiat. 97*:214–225.

Ryans, D. G., 1938a. An experimental attempt to analyze persistent behavior: I. Measuring traits presumed to involve "persistence," *J. Gen. Psychol., 19*:333–353.

Ryans, D. G., 1938b. An experimental attempt to analyze persistent behavior. II. A persistence test, *J. Gen. Psychol., 19*:355–371.

Sanford, R. N., 1936. The effects of abstinence from food upon imaginal processes: a preliminary experiment, *J. Psychol., 2*:129–136.

Sanford, R. N., 1937. The effects of abstinence from food upon imaginal processes: a further experiment, *J. Psychol., 3*:145–159.

Siipola, E. M., 1935. *A group study of some effects of preparatory set,* Psychol. Monogr., 46, No. 210.

Thomas, W. F., and Young, P. T., 1942. A study of organic set: immediate reproduction, by different muscle groups, of patterns presented by successive visual flashes, *J. Exper. Psychol., 30*:347–367.

Thornton, G. R., 1939. *A factor analysis of tests designed to measure persistence,* Psychol. Monogr., 51, No. 229.

Thornton, G. R., 1940. How general is the factor of "persistence"? A reexamination and evaluation of Ryans' results, *J. Gen. Psychol., 23*:185–189.

Travis, R. C., and Kennedy, J. L., 1947. Prediction and automatic control of alertness. I. Control of lookout alertness, *J. Comp. Physiol. Psychol., 40*:457–461.

Warden, C. J., *et al.*, 1931. *Animal motivation,* New York, Columbia University Press.

Wertheimer, M., 1923. Untersuchungen zur Lehre von der Gestalt, *Psychol. Forsch., 4:*301–350; also in *A source book of Gestalt psychology,* ed. W. D. Ellis, New York, Harcourt, Brace & Co., 1938.

Wittenborn, J. R., 1943. Factorial equations for tests of attention, *Psychometrika, 8:*19–35.

Woodrow, H., 1914. *The measurement of attention,* Psychol. Monogr., 17, No. 76.

Woodrow, H., 1916. The faculty of attention, *J. Exper. Psychol., 1:*285–318.

Woodworth, R. S., 1937. Situation-and-goal set, *Am. J. Psychol., 50:*130–140.

Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt & Co.

Young, C. W., and Supa, M., 1941. Mnemic inhibition as a factor in the limitation of the memory span, *Am. J. Psychol., 54:*546–552.

Chapter 4

Perception and Thought

Seeing is believing.

THE language of thought contains many metaphors taken from the language of perception. "Insight" and "illumination" are familiar illustrations, and "I see" is often used to mean "I understand" or even "I agree." Other terms used in the same way are point of view, frame of reference, gap, and find.

The word "perception" itself, in addition to its principal meaning of obtaining knowledge through the senses, has a secondary meaning of cognition or understanding. Such multiple meanings have always caused trouble in psychology, and this extended meaning of perception is particularly likely to produce confusion because principles that hold for one meaning of the word may or may not hold for the other. Therefore in this book we shall be careful to use perception and related words only in their restricted meaning. We shall say "perception of relations" when the relations are perceived directly by the senses. When the relations are abstract, such terms, as "cognition," "knowledge of relations," and "understanding" are preferable.

The popularity of these figures of speech is not due to accident or carelessness alone. Perception and thought overlap in at least three ways:

1. If thought means what people think about or attend to, certainly our thoughts are often thoughts of perceived objects.

2. Thought also means problem solving and judgment. The problem is often a spatial one involving perceived relations, as in working mechanical puzzles and jigsaw puzzles, judging areas of rectangles, designing a dress, and the like. To understand the solution of prob-

lems of this type we must understand the perceptual processes by which the materials of thought are produced.

3. Regardless of materials or contents, the processes of thought and the processes of perception have much in common. Among others Binet (*1903*) drew attention to this functional overlap many years ago, and more recently Gestalt psychologists have attacked the psychology of thought with principles of organization derived from the psychology of perception (Wertheimer, *1923, 1945;* Koffka, *1935*). Just as some psychologists have taken exploratory activity as their model for problem solving, Gestalt psychologists have taken perception for theirs. Both models are helpful, of course, and a well-rounded treatment of thinking will make the most of both, without being blinded by the one or led astray by the other.

Methodologically the most important consideration is that we cannot directly observe anyone's perception as we can his action. We have to get at perception indirectly, by asking the subject to describe, draw, or otherwise reproduce what he perceives. We can ask him to match what he perceives with a given stimulus in some respect. Or we can arrange conditions so that his responses will permit inferences about his perception. There is nothing scientifically peculiar about thus proceeding indirectly, and these methods have yielded much useful information about perception; but when the evidence becomes confusing, the indirect nature of the evidence must be considered.

INFLUENCE OF MOTIVES AND SETS

We have already noted, in the previous chapter, how perception is tuned up and guided by the set of the moment, whether the set is established by a motive, by instructions, or by a habit. Even when the perceptual task is quite clear-cut, as in the judgment of lines and weights, training with monetary reward and punishment will slant the judgments made on a later occasion (Proshansky and Murphy, *1942*). The figure-ground relationship under certain conditions may be shifted by monetary rewards (Shafer and Murphy, *1943*). Stamps and coins are perceived as larger than valueless objects of the same objective size (Ansbacher, *1937;* Bruner and Goodman, *1947*), especially so by poor children.

In this respect perception is no different from other functions of the organism in its environment, being jointly determined by organismic and environmental factors. Certainly a motivationally controlled per-

ceptual apparatus has a survival value for any complex mobile organism, especially if the pattern of motivation includes a natural tendency to react to objects in the environment (Woodworth, *1947*). The detailed relationships between motivation and perception are still to be worked out, but the general principle is beyond question. We can turn therefore to the influence on perception of stimuli in the environment. Then we shall note some more complicated cases, of particular importance for thought and judgment, in which both organismic and environmental determinants are interwoven.

PERCEPTUAL ORGANIZATION AND REORGANIZATION

The important independent variables in the stimulus field seem to be gradients of stimulation, contours, and patterns, and the psychology of perception from the stimulus side consists largely in the analysis of the stimulus conditions that produce patterns: why one pattern is perceived rather than another, why some parts of a complex pattern are perceived more prominently than others, how a pattern may be broken up and the parts reorganized into other patterns, and the like. We are not interested in perception for its own sake in this chapter, to be sure, but merely in those concepts and principles that will be useful in later chapters.

Wertheimer (*1923*) pointed out some years ago how the patterns we perceive are determined by such factors as similarity and proximity. Illustrations of the effects of these Gestalt factors may be found in many textbooks. The importance of such factors in the production of concepts and other patterns will be considered in a later chapter.

EMBEDDING

As a general principle the whole determines the part. The parts of a well-formed pattern are embedded in that pattern and are not readily perceived separately. Though it is difficult to state the requirements for a well-formed pattern, or good figure—aside from the factors of similarity and proximity—with any generality and without circularity, it is easy to show how a part pattern can be embedded and thus effectively concealed in a larger pattern.

In Figure 4.1, taken from Gottschaldt (*1926*), the patterns on the left are fairly well concealed in the patterns on the right. Mere com-

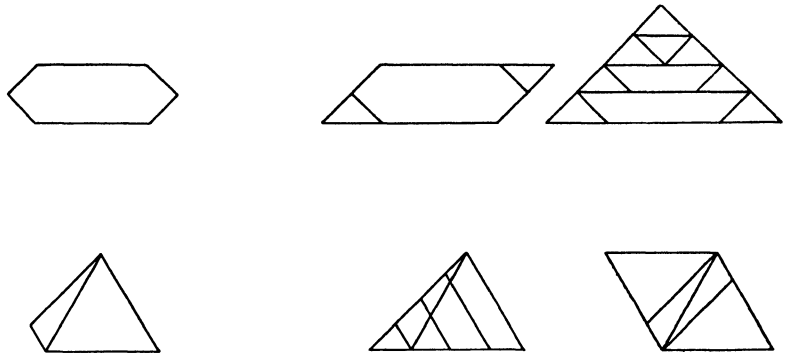


FIGURE 4.1. Embedding. In each row the figure on the left is concealed in each of the figures on the right. After Gottschaldt (*1926*).

plication is not the same as embedding, for the complications of the patterns in Figure 4.2 do not hide the original patterns at all.

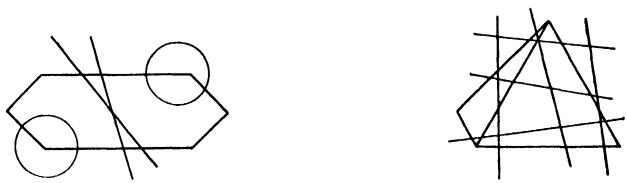


FIGURE 4.2. Complication without embedding.

The principle of embedding may also be applied to the perception of typographical patterns. It is moderately difficult, for example, to find an animal word in the following line:

A Robe A Robe A Robe

It is easier to perceive an animal word in this line:

ZYZABEARSTUVW

And it is very easy to perceive it in this line:

XYZ BEAR stuvy

This principle of embedding is of central importance for the psychology of problem solving, and will be very useful in later chapters.

PERCEPTUAL REFERENCES

Perception is usually focalized, momentarily at least, one pattern, called the figure, being clearly perceived against a background that is vaguely perceived. Under these conditions the background definitely influences perception of the figure, but the background itself is likely to make a weak impression and to be quickly forgotten. If the perceiver is asked to describe what he perceives, he will usually describe the figure and ignore the ground.

Some of the conditions that make part of the stimulus field the ground and part the figure are well known. For example, a portion of the stimulus field that is large and homogeneous is likely to be perceived as ground, while a portion that is small and sharply differentiated from the ground is likely to be perceived as figure. If someone is asked to describe Figure 4.3, he will in all probability describe the X and ignore the background. If one were to reach out and grasp any portion of this stimulus field, he would grasp the X rather than the space around it.

FIGURE 4.3.

REFERENCE POINTS. A case of some importance for the psychology of judgment is illustrated in Figures 4.4 and 4.5. The perceived location of the X will be influenced by its relation to a salient object in the field, such as the O. It will be perceived as left or right depending upon its relation to this prominent object. Such an object can be called a point of reference, or reference point, object, or stimulus.

FIGURE 4.4.

FIGURE 4.5.

In cases like Figures 4.4 and 4.5 the stimulus field is ambiguous and either the circle or the cross could be the reference point for locating the other. The perceiver may impose structure on such an ambiguous field by arbitrarily selecting either one as the reference point. Or an experimenter may instruct his subject to regard one as the reference point. In psychophysics when two stimuli, exposed at approximately the same time, are to be compared, one is designated as the standard or reference stimulus, and the other is then called the comparison or variable stimulus. The organization of the stimulus field is thus determined in part by a set established by social interaction as noted in the previous chapter.

REFERENCE LEVELS. Many objects are perceived in reference to a line, or plane, or level. The X in Figure 4.6 is high; the X in Figure 4.7 is low. The influence of the floor, ceiling, horizon, and the like in the perception of objects is obvious. Reference level is a convenient term in these cases, and in auditory perception. Reference levels and reference points may also be called norms.

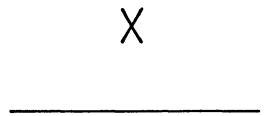


FIGURE 4.6.

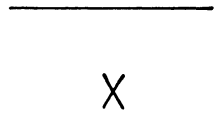


FIGURE 4.7.

REFERENCE SCALES. The scale of reference can best be illustrated in respect to the perception of size. In reference to the five X's of Figure 4.8, the underlined X looks rather small. But the same X looks rather large in reference to the five X's of Figure 4.9. The world is not the

FIGURE 4.8.

FIGURE 4.9.

effective background, nor even the page, but the particular scale to which this particular object is referred. The scale is a one-dimensional, linear pattern organized out of the five objects according to the principles of similarity and proximity.

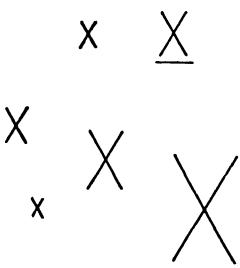


FIGURE 4.10.

Figure 4.10 illustrates a slightly more complex case. The perceived objects are not organized in a scale. Some reorganization is necessary before a scale is formed as a reference for perception of the underlined X.

When the stimulus field is still more complicated, as in Figure 4.11, the perceiver's task is more complicated. He abstracts from the stimulus field the dimension of size or, in other words, he adopts a set to perceive size

and ignore irrelevant aspects such as position, brightness, meaning, and shape. This may be a feat of considerable difficulty and may not be as direct an operation as is implied by the term perception.

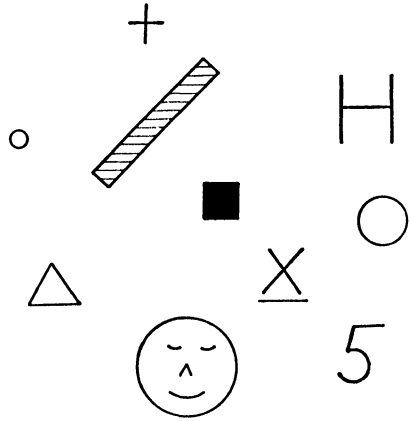


FIGURE 4.11.

When the objects in view vary in more than one dimension, maintenance of a set for one dimension rather than another is difficult. In fact the effect of interference or retroactive inhibition in psychophysical judgments of this sort been clearly demonstrated by Postman and Page (*1947*). They had their subjects judge a series of rectangles for height; then judge a series for width, and again for height. The precision of the second series of height judgments was significantly disturbed by the interpolated activity, no doubt because of the interference of the set to perceive width on the set to perceive height (see Table 4.1).

Perception of the variegated objects of the real world is often complicated by additional intellectual operations, and in such cases the response is less direct, less certain, and less rapid. These additional operations are commonly treated under the heading of judgment. The organization of a scale of reference, for example, has come in for considerable quantitative investigation, which will be reviewed in Chapter 10.

FRAME OF REFERENCE. Frame of reference is another term that strayed from the psychology of perception into the neighboring fields of social psychology and the psychology of judgment. In some respects

TABLE 4.1. Effects of Retroactive Inhibition on Set for Judgments of Rectangles[a]

Series	*h*
1st training	.12
2nd training	.16
3rd training	.22
1st interpolated	.13
2nd interpolated	.17
3rd interpolated	.21
Test	.11

[a] Three subjects judged widths, then heights, then widths. Three judged heights, then widths, then heights. *h* is a measure of precision of judgments. Extracted from Postman and Page (*1947*).

frame of reference has the same meaning as a scale of reference. Sherif and Cantril (*1947*), for example, in a chapter on "Experimental Factors Concerning the Basis of Attitudes," use both terms, scale and frame. Scale of reference is the more precise term and will be used in this book when the object of perception or judgment is referred to a one-dimensional pattern, i.e., to a scale.

The chief source of confusion is the use of frame of reference to include everything that influences perception or judgment or attitude. The important contribution of the concept of frame of reference is almost the opposite. Perception does not depend upon everything. The object of perception, or judgment, is referred, not to the whole world, but to a specific background or framework. The experimenter cannot construct two different worlds, but he can construct two different frames of reference within which a constant object will be differently perceived.

Furthermore, the frame of reference, like the scale of reference, is not just another specific item, or twenty more items. It is a structure, a pattern, a framework. It is important, if this term is to retain its usefulness for psychology, to restrict it to backgrounds or frameworks, segregated out of the environment, to which an object of perception or judgment is referred.

The frame of reference may also be confused with the set. This is understandable because, as noted above, the effective scale or frame may be determined by the perceiver's set. If he is judging size, he abstracts the size dimension from other dimensions, and the perceived sizes then constitute the background to which any particular size is referred. But

the two concepts are not the same. The set may be experimentally varied by instructions, while the frame of reference may be varied by changing the objects in view.

The most precise and indisputable meaning of frame of reference may be illustrated by Figures 4.12 and 4.13. Obviously, whether the X is perceived as right or left, high or low, depends upon the background or frame of reference within which it is perceived. If the figure does

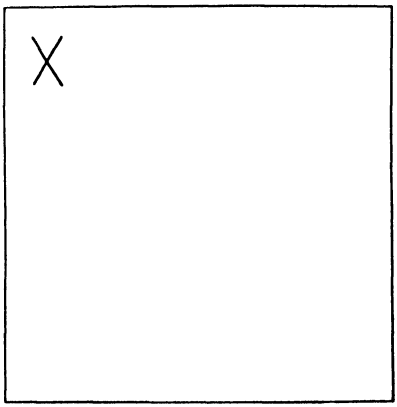


FIGURE 4.12.

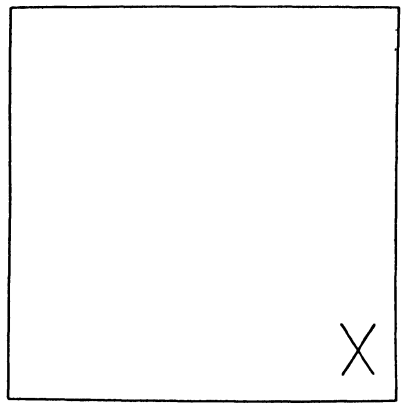


FIGURE 4.13.

not have a definite frame of reference, as when a point of light is seen in darkness, localization is unstable. This is the sense in which the term was used by Koffka (*1922, 1935*), and by Sherif (*1935*) in reporting his experiments on the autokinetic effect. In this sense the term is unequivocal and, at the same time, very useful. This is the usage that should be preserved when the term is applied to the psychology of judgment.

UTILITY OF PERCEPTUAL CONCEPTS FOR THOUGHT AND JUDGMENT

The perceptual concepts introduced in this chapter are the concepts of most promise for the investigation of thought and judgment. We may accept these concepts as potentially useful without asserting that judgment (or problem solving) is the same as perception. Chapter 2 attempted to describe in general how psychological functions are to be differentiated and Chapter 9 will discuss the relations between perception and judgment in particular.

Nor must we say that the laws of perception are the laws of thought. Dynamic principles, derived from experiments on reaction time, muscle tension, and the like, will be useful in the study of thought and judgment, and so will principles of learning, to be reviewed in the next chapter. In the same contributory way principles of organization, derived from experiments on perception, will be useful, but not sufficient alone.

INDIVIDUAL DIFFERENCES

Perceptual tasks have supplied the materials for many tests of individual differences in perceptual abilities, some of which are important for thought and judgment.

PERCEPTUAL ABILITIES

It has long been known that tests which require the manipulation of visually perceived patterns correlate well with each other. Tests that call for matching patterns, counting blocks, and the like are in common use as homogeneous tests and as items in tests of general intelligence. Thurstone's (*1938*) factor analysis of 56 tests, for example, disclosed a visual space factor, and his Primary Mental Abilities battery (Thurstone and Thurstone, *1943*) includes tests of this ability.

Many perceptual tests of various types were tried out by the Aviation Psychology Program of the Army Air Forces (Guilford, *1946*). Some of these were analyzed factorially. The usual perceptual speed factor was identified and found useful in the selection of pilots and navigators. A useful spatial relations factor and a useful visualization factor also turned up, each of which has something in common with Thurstone's perceptual factors. An interesting, rather clear-cut factor of length estimation appeared. The Gottschaldt figures came out in a secondary reasoning factor and in the visualization factor.

The factor-analysis approach has not gone far and most of the results need confirmation with different tests and different subjects. But some support has already been added to the assumption that principles of perceptual organization may be applied to problem solving in other areas, e.g., verbal reasoning.

PERCEPTUAL SETS

Aside from variations in abilities, different people are likely to perceive different things in the same environment simply because they

are looking for different things. One may be looking north while the other is looking south. One may be looking for gold while the other is looking for a place to sit down.

Such sets come and go, of course, but people differ also in more permanent ways of orienting to the perceived world. Every mobile organism is probably set, when awake at least, to perceive objects in the environment, especially obstacles and pathways, but some are slanted more than others toward some objects or aspects of the environment. When children meet strangers, they notice superficialities like height and age, while adults are more likely to notice signs of occupation and social status. Children up to about the sixth year match objects on the basis of color rather than shape (Brian and Goodenough, *1929*). Any good observer can notice how people with different attitudes toward the world will perceive different things in it.

Postman, Bruner, and McGinnies (*1948*) have devised a technique for showing in the laboratory how differences in motivation produce differences in what is perceived. They were concerned, not with primary biological drives, but with personal values, as measured by the Allport-Vernon Study of Values: theoretical, economic, aesthetic, social, political, and religious values. They made up lists of words to represent each of these values: theory, verify, sciences, etc., for the theoretical value; income, useful, wealthy, etc., for the economic value; beauty, artist, poetry, etc., for the aesthetic value; and similar words for the others. Each word was then exposed tachistoscopically to a subject three times for .01 second. If the subject failed to recognize the word, three exposures were then given at .02, .03 second, etc., at exposure times increasing in even steps of .01 second until recognition occurred. A full record was kept of all the subject's prerecognition responses. The subjects were instructed simply to report everything they saw or thought they saw.

Analysis of a typical subject's recognition times showed that if he had a high score for the theoretical value and a low score for the economic value, his average recognition time for words in the theoretical category would be relatively short and for the economic category relatively long. Throwing all subjects together, words in the category of first value rank were perceived quickly on the average, words of the second value rank less quickly, and so on.

There was a clear-cut relation for most subjects taken individually and for the group as a whole between their values and their speed of

perception. The same effect has been demonstrated when the words are presented aurally (Vanderplas and Blake *1949*).

Another interpretation should also be considered. It is possible that the dynamic effect of personal values operates prior to the experiment to increase the availability of words of high value. A person who values economic matters highly is likely to use words in the economic category frequently and willingly, to have them conveniently and facilely on tap. If this interpretation is correct, words in a category of high value would appear more often in an anagram problem as well as a perceptual problem. There are other interpretations, but it seems clear that individual differences in personal values will affect people's activities in any problematic situation that contains objects of different value. There is further evidence that people who have been experimentally subjected to failure will perceive different objects in an ambiguous stimulus field than people who have been allowed to succeed. (McClelland, Atkinson, and Clark, *1949*).

PERCEPTUAL REFERENCES

Individual differences in perception are produced when different observers perceive the same object against different backgrounds. Trained weight lifters, used to lifting bar bells, judge laboratory weights lighter than watchmakers do, no doubt because they are accustomed to different scales of reference (Tresselt, *1948*). A man who measures six feet in height looks tall to most people but not to a basketball coach. This obvious principle will be very helpful in later chapters on the psychology of judgment.

REFERENCES

Ansbacher, H., 1937. *Perception of number as affected by the monetary value of the objects,* Arch. Psychol., No. 215.

Binet, A., 1903. *L'étude expérimentale de l'intelligence,* Paris, Schleicher Frères.

Brian, C. R., and Goodenough, F. L., 1929. The relative potency of color and form perception at different ages, *J. Exper. Psychol., 12:*197–213.

Bruner, J. S., and Goodman, C. C., 1947. Value and need as organizing factors in perception, *J. Abnorm. & Social Psychol., 42:*33–44.

Ellis, W. D., 1938. *A source book of Gestalt psychology,* New York, Harcourt, Brace & Co.

Gottschaldt, K., 1926. Uber den Einfluss der Erfahrung auf die Wahrnehmung von Figuren: I. Uber den Einfluss gehaufter Einprägung von

Figuren auf ihre Sichtbarkeit in umfassenden Konfigurationen, *Psychol. Forsch., 8:*261–317; also in Ellis, above.

Koffka, K., 1922. Perception: an introduction to the Gestalt-theorie, *Psychol. Bull., 19.*531–585.

Koffka, K., 1935. *Principles of Gestalt psychology,* New York, Harcourt, Brace & Co.

McClelland, D. C., Atkinson, J. W., and Clark, R. A., 1949. The projective expression of needs: III. The effect of ego-involvement, success, and failure on perception, *J. Psychol., 27:*311–330.

Postman, L., Bruner, J. S., and McGinnies, E., 1948. Personal values as selective factors in perception, *J. Abnorm. & Social Psychol., 43:*142–154.

Postman, L., and Page, R., 1947. Retroactive inhibition and psychophysical judgment, *Am. J. Psychol., 60:*367–377.

Proshansky, H., and Murphy. G., 1942. The effects of reward and punishment on perception, *J. Psychol., 13:*295–305.

Shafer, R., and Murphy, G., 1943. The role of autism in a visual figure-ground relationship, *J. Exper. Psychol., 32:*335–343.

Sherif, M., 1935. *A study of some social factors in perception,* Arch. Psychol., No. 187.

Sherif, M., and Cantril, H., 1947. *The psychology of ego-involvements,* New York, John Wiley & Sons.

Thurstone, L. L., 1938. *Primary mental abilities,* Psychometric Monogr., No. 1, Chicago, University of Chicago Press.

Thurstone, L. L., and Thurstone, T. G., 1943. *The Chicago Tests of Primary Mental Abilities,* Chicago, Science Research Associates.

Tresselt, M. E., 1948. The effect of the experiences of contrasted groups upon the formation of a new scale of judgment, *J. Social Psychol., 27:*209–216.

Vanderplas, J. M., and Blake, R. R., 1949. Selective sensitization in auditory perception, *J. Personality, 18:*252–266.

Wertheimer, M., 1923. Untersuchungen zur Lehre von der Gestalt, II, *Psychol. Forsch., 4:*301–350; also in Ellis, above.

Wertheimer, M., 1945. *Productive thinking,* New York, Harper & Brothers.

Woodworth, R. S., 1947. Reinforcement of perception, *Am. J. Psychol., 60:*119–124.

Chapter 5

Learning and Thinking

> If I saw farther, 'twas because I stood on giant shoulders.
>
> —NEWTON

BOTH learning and thinking are ancient concepts, with long histories of speculation and investigation. This chapter aims to clarify the relations between the two concepts and to summarize the contributions of learning to the psychology of thought and judgment. For this purpose three methods outlined in the chapter on methods of analysis will be used: logical analysis, experimental analysis, and analysis of individual differences.

LOGICAL ANALYSIS

If the analysis could be limited to the behavior of a very simple animal in a very simple situation, it might be sufficient to treat practice as another independent variable, to be directly related to the dependent variables of behavior. But man is a very complex and flexible animal, and this book must consider the most complex things he does. The important thing about practice is that its consequences endure so long and influence subsequent behavior at so many different times and in so many different ways. Of a dozen facts learned in a schoolroom some may later be used to solve a problem in that schoolroom, but some may be used to solve a problem at another place and time, and some may be combined with other facts learned elsewhere to solve still different problems at different places and times. Hence it is scientifically fruitful, instead of studying the relation between problem-solving behavior and previous experience directly, to interpose the consequences

of previous experience, such as vocabulary, information, and knowledge.

This system of concepts is an ancient one, implicit in popular usage and in much psychological research. But before the contributions of research on learning to our topic can be summarized, some clarification of learning and related concepts is necessary.

A BROAD VIEW OF LEARNING

When we say that man is an adaptable animal, with great capacity to learn, we are speaking of one of two principles of development, the other being maturation. Our concept of learning must be broad, for between learning and maturation all modifications of behavior, including perceptual-motor skills, language, knowledge, attitudes, and personality traits, must be accounted for. Sophisticated distinctions between these two concepts have been attempted, but it is enough for present purposes to remember that maturation is a phylogenetic sort of adaptation, practically independent of the postnatal life history of the individual, while learning accounts for the adaptation of a particular individual to its particular environment. Although the two principles of development are conceptually independent in this sense, in another sense they are closely related because the individual's learning capacity is dependent upon its maturational status at any moment in its life span.

Although learning is, in a broad biological sense, an adaptation to the environment, in specific instances we cannot always say that the consequences of learning are adaptive, if we mean by this that they further the organism's vital activities. Therefore, when we come to close grips with the matter, we speak of learning merely as producing changes in activity, leaving the evaluation of those changes an empirical question.

Not all changes in activity are properly subsumed under the learning process. Sophisticated definitions of learning find it necessary to rule out changes due to maturation, fatigue, sensory adaptation, accommodation, and the like. The changes due to learning are dependent upon practice, training, or experience. They are more or less lasting and are related in some way to the organism's motivation or general activity in progress.

For the sake of completeness such activities as imagination and judgment should be mentioned here. Suppose someone looking out a window imagines that he is on the deck of a three-masted schooner

looking over a boiling sea. No honest psychologist can assert that such activities do not occur or that psychology should ignore them. Or suppose that someone looking out the window judges the distance to the ground below him. At first it looked like six feet, but now he estimates it as ten feet. These activities of imagination and judgment change something, but can such changes be listed under the heading of learning when learning is broadly conceived? The criterion of duration seems to supply the answer. When these activities, like fatigue and sensory adaptation, have no lasting consequences, they do not contribute to the development of the individual, and the psychology of learning can ignore them. But when the results of imagination, judgment, and the like are retained and affect subsequent behavior, a broad concept of learning must include them.

Within this conceptual framework empirical investigation of learning proceeds. Since by definition a learning experiment involves a comparison of activities on two occasions, with a finite period of time in between, the topic is conveniently divided into three phases: (1) the performance of new or different activities, (2) retention of what is new, and (3) testing for the effects of these previous activities on later activities, such as recall and recognition, decrease in errors, qualitative changes, and the like. The accompanying chart is intended to show these relationships.

A System of Concepts Related to Learning

I. Maturation

II. Learning, experimentally analyzed into three phases:
   - A. Acquisition of new activities, by
     1. conditioning
     2. variations in activity with motivational reinforcement, i.e., trial and error
     3. imitation, tuition, reaction formation, and other social processes
     4. thought, reasoning, reorganization and insight, creative imagination, judgment
     5. et cetera
   - B. Retention
   - C. Consequences on later occasions, such as
     1. recall, recognition
     2. repetition of new activities

Experimentally, one phase of learning is isolated for study by the classical procedure of holding the other phases constant. Research on the first phase centers around a description of the processes by which the new activities are acquired. It is altogether logical to include thought, reasoning, and the like under this heading, along with simpler processes.

McGeoch (*1942*) believed that learning consists typically of a problem situation and that the first phase is the "discovery of an adequate solution." There are instances of learning in which the problem situation and the discovery of its solution seem to be minimal, but so much of learning clearly contains them that they must be regarded as typical of at least a wide range of phenomena, and they may be interpreted as typical of all." Nevertheless the term "discovery of an adequate solution" seems too strong to include such processes as conditioning and imitation. In this book, not to put too much of a load on interpretation, we shall use a more noncommittal term, such as "initial performance of the new activity" or "acquisition phase."

In a broad conception of learning, then, the topic of thought and judgment finds its place as one way of acquiring new activities. The question whether the new activities are retained and reproduced on a later occasion belongs to the psychology of learning.

Research on the third phase compares the different methods of obtaining evidence that new activities have been acquired and retained, and assays the effects of such variables as motivation and distraction on recall, reperformance, relearning, and the like. When anyone is asked to give evidence that he has mastered a poem or a puzzle, he may run through it rapidly and smoothly. He may, however, have to "stop and think." Remembering may be selective and constructive as well as reconstructive. Imagination may be productive as well as reproductive. And if two memories are equally strong, an act of judgment may be required. Therefore it appears that, in a broad view of learning, thought and judgment may be included under the recall or memory phase as well as under the acquisition phase. Or, to state the matter systematically, thinking may take place in any problematic situation.

This way of arranging the concepts in this area of psychology is intended merely to make explicit and consistent the usage of much of modern psychology. Melton's (*1941*) arrangement is similar. Hilgard (*1948*) likewise puts problem solving, reasoning, and the like under the heading of learning: "Leaving the doubtful processes in simply

asserts that a complete theory of learning must have something to say about reasoning, creative imagination, and inventiveness, in addition to what may be said about memorizing and retaining or about the acquisition of skill."

A NARROW VIEW OF LEARNING

Unfortunately the term learning is also used in a narrow sense, by a kind of unintentional synecdoche, to refer only to the acquisition of new activities by the simpler processes. Learning then is contrasted with thinking, problem solving, and reasoning.

Maier's (*1931*) handling of the concepts in this area of psychology appears at first glance to conform to this narrow view, for he explicitly separates learning from reasoning. But in a systematic book on animal psychology, Maier and Schnierla (*1935*) really take a broad view of this area, using a different arrangement of concepts. What in the present treatment is called learning, Maier and Schnierla call modification of behavior. Under that heading they include conditioning, learning in the narrow sense exemplified by simple maze learning, and reasoning. Retention, recall, and relearning in this scheme would presumably not be considered under learning but under modification of behavior. A priori, one system of concepts is as good as another, but the system set forth in the above chart comes closer to the usage of modern psychology.

Fortunately these differences in usage do not cause as much confusion as one might expect. Many writers who contrast learning with thinking, reasoning, or inference add a qualifying adjective and call it simple learning or associative learning. The important thing about a concept is not its name but where it fits into a larger system of concepts.

ANALYSIS OF THE ACQUISITION PHASE

Within the more comprehensive system of concepts illustrated in the chart above, it becomes necessary to differentiate thought from other ways of initiating or acquiring new activities. This is often done, quite casually, by calling thought, creative imagination, reasoning, and the like higher or rational forms of learning—meaning the acquisition phase of learning—to be contrasted with lower or simpler forms of learning, such as conditioning, imitation, and the memorizing of associations.

Use of the terms simple and complex, or higher, forms of acquisition has the advantage of suggesting a continuum rather than distinct categories. Furthermore these are superficial, noncommittal terms that correspond better to the present vague state of analysis of the acquisition phase than more sophisticated but perhaps misleading terms.

DIMENSIONAL ANALYSIS

McGeoch (*1942*) and Melton (*1941*) go beyond this casual distinction and recommend a dimensional analysis of the acquisition phase of learning. The different ways of initiating a new performance, they say, can and should be analyzed to show up similarities and differences in materials and processes. One dimension, for example, is the degree to which the learner has to discover the adequate response. He may be given the material with instructions to master it, or he may be required first to discover a solution, then fixate it. Other suggested dimensions are the motor as contrasted with the ideational or verbal, the incidental as contrasted with the intentional, and the degree of relationship among the materials to be learned. Another candidate for inclusion is the social dimension. The learning situations of real life and even of the laboratory differ significantly in the extent to which social variables influence acquisition of new activities.

This is a peaceful and therefore valuable approach, but, since it does not have the dramatic appeal of those approaches that interpret all learning in terms of a favored prototype, it has not been followed up to any extent. No one, with the possible exception of Welch (*1947*), has actually arranged a variety of instances of learning along any dimension. Two or three instances do not make a dimension. The recent appearance of instrumental conditioning, however, lying between classical conditioning and problem solving on the dimension of discovery of the adequate response, does illustrate the systematic potentialities of this approach. Other possibilities have been noted by Underwood (*1949*).

In any event, if a number of people were observed all day for several days, so that a large representative sample of instances of acquisition of new activities could be gathered, it would be worth while to try to sort these instances and locate them along several dimensions. Surely the dimensions would not be orthogonal, for the degree of relationship among the materials to be learned, for example, is not

independent of the degree of discovery involved. Whatever the outcome of dimensional analysis, it is obvious that thought, reasoning, and the like are at one extreme of the discovery dimension and at the intentional end of the incidental-intentional dimension. The location of these higher forms of acquisition on the other dimensions, e.g., the motor-verbal dimension, would vary considerably.

The principal shortcoming of dimensional analysis to date is that no method of analysis, other than inspection, has been described. If we really want a method of finding similarities and differences in a sample of learning tasks, analysis of individual differences may be suggested. Those tasks are similar which the same people do well. Actually this method of analysis is being used, and some of the available evidence will be summarized toward the end of this chapter. In general this evidence supports the rough-and-ready distinction between simple and higher forms of learning.

EXPERIMENTAL ANALYSIS

Maier's (*1931, 1940*) distinction between learning and reasoning has been supported by ingenious experiments on animal and human subjects. By learning he means the acquisition of new activities by simple association, and by reasoning he means a higher kind of acquisition, spontaneous integration of separate experiences.

ANIMAL SUBJECTS. In a typical experiment Maier (*1929*) allows a rat to explore the experimental situation and gain general familiarity with three tables and paths from each to the others (Experience 1). Then the rat is fed on one of the tables (Experience 2). When it has been put on one of the other tables, running directly to the food table could be interpreted, not as simple associative learning, for it has not previously taken that path to food, but as spontaneous integration of Experience 1 and Experience 2, i.e., as reasoning.

Whether what rats do on Maier's three-table problem should be called reasoning or not is subject to some dispute. Hull (*1935*), without discussing the name, has reinterpreted such performance in terms of fractional anticipatory goal responses. Seward (*1949*) devised an experiment aimed at a theory of latent learning which turned out to be very similar in design and results to Maier's reasoning experiment. As pointed out in Chapter 2, the separation of something from something else does not convincingly identify either.

But regardless of the name for what the rats do or the interpretation, Maier's surgical separation of the two activities must be taken seriously. Admitting that reasoning and learning usually occur together, he set up a way of scoring successes due to R + L and also, by contrasting the consequences of each, to R — L. Then, by a little algebra, achievement due to R alone and to L alone can be computed. This scoring procedure was applied to the performance of a group of rats before and after brain surgery. The result was that the cortical lesions (up to 41 percent) affected reasoning quite definitely, but not learning. Whether the capacity called for by the three-table problem be called reasoning, ability to maintain an orientation, place learning, capacity for anticipation, or what not, it apparently is different from simple learning in that the former is affected by the lesions while the latter is not.

Varying the Working Conditions. In his work on the reasoning of human subjects Maier has not been able to make the same surgical dissection of problem solving but has had to proceed indirectly. The strategy is the same, however. Reasoning or productive thinking is the higher way of acquiring new response patterns. Mechanical response to previously learned stimuli or equivalent stimuli is the simpler way.

The latest experiment (1945) investigates the role of equivalent stimuli and the adequacy of this concept to account for productive thought or reasoning. The hypothesis is set up that if what has been learned in the past can be transferred to the present problem (reproductive thinking) and utilized in the form in which it was first learned, no productive thinking would be necessary. To test this hypothesis Maier used his well-known hatrack problem. The problem is to construct a hatrack given only limited material, including two sticks and a clamp. The previous learning was the solving of a problem, aided by the experimenter if necessary, in which sticks were clamped together to support a string pendulum from the ceiling. The critical solution to the hatrack problem is to clamp the sticks together as a pillar between floor and ceiling and hang the hat on the clamp. According to Maier, the interpretation in terms of equivalent stimuli is that experience with the string problem would be carried over to the hatrack problem and solution would be immediate, i.e., no productive thinking would be necessary.

Three groups of 25 each attempted the hatrack problem. One group had experience with the string problem and worked on the hatrack

problem with the finished string problem in view. Another group had experience with the string problem but was removed from sight of the string structure. The third had no experience with the string problem. It turned out that 18 students in the first group solved the problem, 12 in the second, and only 6 in the third.

Several conclusions are possible. In the first place the experiment shows that previous experience with the string problem was not necessary, for about a fourth of the last group solved the hatrack problem. But experience with the string problem was helpful because both the first and second groups did much better than the last. And furthermore a comparison of the first and second groups shows that experience with and continued exposure to the string problem were more helpful than experience alone.

The question now arises: How did continued exposure to the string structure increase the number of solutions to the hatrack problem? Maier's interpretation is that equivalence of stimuli, though helpful, is not sufficient. Those subjects to whom the stimuli were not equivalent had to reorganize their experiences, aided by a view of the string structure. But it is really very difficult to arrange an experiment that will differentiate recall of past experience from reorganization of experience. One could argue that having the string structure in view merely aided recall of what had been learned shortly before. And the time data in this case are not helpful because there is no way to determine whether reorganization of experience or recall of equivalent stimuli requires more time.

There is an important methodological question implicit in this kind of experimentation. The fact that someone sees a tool used for one purpose or even that he helps make the tool cannot be taken as proof that he has learned the use of the tool for that purpose any more than the fact that he sees a flight of stairs or even ascends them one at a time proves that he has learned the number of steps in the flight. In a sense the question is one of the criterion of mastery, of the test to be used in the third or recall phase of learning to demonstrate what has been acquired and retained. It is relatively easy to set up a criterion of mastery if the experimenter knows what the subject is trying to master. But if, as in Maier's experiments, the experimenter knows only what the subject has been exposed to, it is very difficult to assay the residue of his exposure. One could, of course, ask the subject to describe his experience introspectively, as was done so often in the early days of

psychological investigation of the thought processes, but it is doubtful that specific valid information would be obtained. Maier, in the experiment just reviewed, has pointed the way to an answer to this troublesome question by exposing some subjects to a general demonstration of the sticks and clamp in finished position and asking them to write down what useful functions the structure could serve. Relatively few—11 percent by a strict classification and 24 percent by a generous classification—reported the function of the structure as anything resembling a hatrack function. These were not the subjects who attempted the hatrack problem, so it is not possible to make direct use of their answers. Maier takes this 24 percent as the maximum effect of equivalence of stimuli and compares it to the 72 percent success of the first experimental group, and concludes that equivalence of stimuli will not account for problem-solving achievement. This is an ingenious use of the subjects' comments but would not convince a skeptic, for the conditions of exposure to the string structure were more favorable to learning the functions of the structure for the active experimental group than for the passive spectators of the last-mentioned group.

VARYING THE INSTRUCTIONS. Maier's method of differentiating two processes depends on two different laboratory arrangements. Another method, commonly used in educational psychology, is to give all subjects the same materials to work with, but to instruct different groups in work in different ways. Then successes and errors of the different groups are compared.

Katona's (*1940*) studies of organizing and memorizing illustrate this method. In one experiment he had college students learn card tricks in different ways. One group was given the exact order of arranging the cards, which were memorized by repetition. Another group watched a demonstration and explanation of the principle of the trick, given only once. In these experiments it is usually impossible to observe differences in the intellectual operations of the two groups. Such differences may be inferred, however, if the groups show objective differences in some aspect of achievement, e.g., if one group is significantly more successful than the other. Katona's memorizing group was distinctly less successful in performing the tricks than the organizing group. Similar experiments were carried out with matchstick puzzles, words, and numbers.

Other experiments following this strategy have been reported (Swenson, Anderson and Stacey, *1949;* Szekely, *1950a 1950b*). The

aim is usually the practical one of improvement of teaching methods rather than the analysis of intellectual operations. In general the more difficult, complex tasks are mastered more effectively by organizing or understanding the material than by memorizing, especially when transfer to new materials is called for. When the materials to be acquired are simple, the differences are small, and occasionally drill has been found to be the more effective method. The arguments that different groups of subjects are functioning in different ways are, first, that they are instructed differently; and, second, that their achievements are different. In some experiments of this kind it can be shown that different kinds of errors are made by the different groups.

Conclusions Thus Far. There are several good logical and empirical reasons for grouping several ways of initiating new activities under one heading, the acquisition phase of learning, if learning has a broad meaning. The simpler acquisition processes, classical conditioning, associative learning, and memorizing, can be superficially distinguished from the higher processes, productive thinking, reasoning, creative imagination, and the like.

Going below the level of superficial description, it is apparent that in the higher forms of learning practice is less important than relationships within the situation. This fact has two pertinent consequences. First, when progress in an activity increases regularly as a function of practice, the activity is often called learning, or simple learning. When amount of practice or exposure is not an important independent variable, the task is often called complex learning, problem solving, or thinking. Second, success in complex learning or problem solving is usually indicated, not by the cumulation of practice effects, convenient for plotting as a learning curve, but in either-or form. The problem is solved under some conditions but not under others. It is solved by some people but not by others.

This differentiation is not at all clear-cut. Practice as an independent variable can never be entirely eliminated, nor can relationships within the material of the problem. The higher forms of acquisition may be considered a special case of the simple fundamental principles of learning, with the addition of a few new variables. Or the simpler forms of acquisition may be considered a special case of problem solving, with some of the important variables artificially eliminated.

Perhaps the best argument for recognizing higher forms of acquisition is the crude fact that contemporary discussions of simple learning

have little in common with the contents of the present book. Evidence on this question, derived from analysis of individual differences, to be reviewed later in this chapter, supports this rough-and-ready distinction between simple and higher ways of acquiring new behavior patterns.

These conclusions have two implications for the present work. First, our topic will be the higher, more complex ways by which new activities are initiated. Simpler acquisition processes can be ignored, and practice is not the most important independent variable. The retention and subsequent use of the new activities can be ignored—unless on a subsequent occasion new problems arise. We are interested in the initiation of the new behavior, the first solution to the problem, rather than the subsequent fate of this behavior pattern. Second, since problem solving often makes use of the results of previous learning, a review of what previous learning makes available is indicated.

THE CONTRIBUTION OF PREVIOUS LEARNING TO THOUGHT AND JUDGMENT

Certainly today's problem solving is often, if not always, influenced by yesterday's learning. A familiar metaphor offers an appropriate framework for making this relationship explicit. Memory may be pictured as a storehouse, stocked by previous learning, from which thought selects the raw material for its operations. We must inventory, in general terms but as clearly as possible, the contents of the storehouse, their arrangement and availability for thought and judgment.

SENSORI-MOTOR COÖRDINATIONS

All organisms have a variety of sensori-motor integrations on tap when they approach a problem. A rat, when it hears a bell ring, may habitually turn left. A human subject, when he sees a hole, may habitually put his finger in it. Evidence for such habits is obtained by presenting the organism with the sensory stimulus under optimum conditions and observing the occurrence, or failure, of the motor response, its latency, speed, strength, etc.

PERCEPTUAL ORGANIZATIONS

When anyone enters a problematic situation, he is equipped with a variety of ways of perceiving the situation. There is no doubt that these perceptual organizations, when pertinent, influence problem-solving

activities. Perception may be organized and reorganized during problem solving if it is a spatial problem, but it is also true that the thinker may have several alternative organizations of the sensory data in reserve before he commences work on the problem.

When we ask for an inventory of what the thinker perceives when he thinks about a perceptual problem, the answer is disappointing. Since the lower animals cannot report what they perceive, their perception must be inferred from their behavior and, since other variables enter in, these inferences are usually debatable. This is the principal methodological reason why the relative contribution of motor learning and perceptual learning in mastery of a complex problem constitutes a lively theoretical controversy. Communicative human subjects can talk about what they perceive, but their free comments are not likely to prove satisfactory. Maier's approach of this problem has been mentioned earlier. He asked his subjects, looking at a hatrack structure, what functions it could serve. People could also be asked, at first sight of a spatial problem, if they have seen such a problem before, what relations they perceive between critical parts, what parts are not functionally related to the remainder, where the prominent contours are located, what alternative configurations of the parts they are familiar with, and the like. If questions can be made precise and all answers can be coded, the thinker's initial supply of pertinent perceptual organizations may be catalogued with some degree of completeness. If questions of this kind are not informative, variations of the sorting test and of the matching test might be tried out. There is no reason a priori why subjective characterizations of the perceptual organizations that a thinker brings to a problem cannot be obtained with satisfactory reliability and related to more objective data.

WORDS

The concept of memory as a storehouse of experience becomes more convincing and experimentally more approachable when we consider the solution of verbal problems and the supply of words that a person brings to these problems. An inventory of the supply is relatively easy, in principle at least. Because even the simplest sensori-motor coördinations and perceptual organizations overlap one another, the task of identification and enumeration of these fruits of experience is almost impossible. Words, on the contrary, are discrete, easily elicited, easily identified and counted, and relatively easily classified. To understand

higher intellectual activities involving verbal habits, the information on the supply of words and their availability should be systematized.

THE SUPPLY OF WORDS. Supply of words is investigated by tests that elicit words without specifying their utility in the solution of any particular problem. A variety of testing devices are in use for this purpose: simple recall, aided recall, multiple-choice tests, free and restricted association, continuous association (or recall), and conditionability. The information obtained by each procedure depends, of course, upon the method used but the underlying facts of supply and availability are approached by several different methods.

THE STRUCTURE OF THE SUPPLY. If, after instructing ordinary adult subjects to respond with the first word they think of, we give them the stimulus word *man,* about half will respond with *woman.* To *dark* about half will respond *light.* Since the instructions are nonspecific, favoring no response word more than any other; we must conclude that for many subjects the word *man* has a particular relationship to the word *woman* and *dark* to *light.* That is, the words in the supply are arranged in some sort of structure or organization which influences their availability.

If different people come to an association experiment with the same set, their responses to the stimulus words yield evidence of differences in the structure of their verbal storehouses. We know, for example, that women organize their experiences in this world along different lines than men. If the stimulus word is *blue,* the response word *dress* is given more often by women than by men; *spectrum* more often by men than by women. This is the basis for one of the Terman-Miles (*1936*) tests for masculinity-femininity. Using 171 words and a multiple-choice modification of the free association procedure, they get a reliability coefficient of .78 and only a 10 percent overlap between the sexes. Goodenough (*1942*) used homonyms and got an even sharper division.

It is clear, then, that the stimulus words are more closely related to some response words than to others, and that these relationships are different for men and women. When the set for the response is, in both sexes, simply to give the first word recalled, the relational structure is probably the chief factor determining these sex differences. When adults are compared with children, however, the set for the task may differ. Children seem more eager to produce an impressive word, for example, hence differences in response words cannot be taken unequivocally as differences in the supply brought to the task.

Further information about the structure of the verbal storehouse comes from Razran's (*1939, 1949a, 1949b*) work on semantic conditioning. Razran uses the salivary response, measuring the secretion of saliva by placing a roll of absorbent cotton, such as dentists use, under the subject's tongue, and weighing it before and after the subject looks at the stimulus word. The original or unconditioned stimulus can be a piece of candy or anything that makes the saliva flow. The stimulus-to-be-conditioned is a word exposed on a screen. One of these words was *style,* s–t–y–l–e. It was conditioned to the salivary response, then other stimulus words were tried out to see how much spread or generalization would occur. The most generalization was to synonyms, 59 percent on the average. That is, the salivary increment in the weight of the cotton roll when the subjects were exposed to the synonyms was 59 percent of what it was when they were exposed to the conditioned stimulus word. For style the synonym was fashion. The next largest amount of generalization was to homophones, words that have the same sound but different meaning. For style the homophone was stile, s–t–i–l–e. Other experiments have shown a little spread due to phonetographic relationships, that is, sound spelling, as between dark and mark. In the language of the free association experiments this would be called a clang relationship.

Others have repeated these experiments, using the galvanic skin response (Riess, *1940*), and it seems clear that such generalization of conditioning does occur, the effects spreading to words that are similar or opposite in meaning, similar in sound, and to nonsense syllables that have merely been associated with a conditioned syllable. Apparently the meaningful or semantic relationship is the most prominent one for college students under the usual experimental conditions.

The structure of verbal relationships may of course be very complicated and may extend over several steps. If someone hears the word *needle,* he may think to himself of *sharp,* then *dull,* then he may respond *moron.* The last word *moron* is not a direct response to *needle* but is the result of an indirect association. Cofer and Foley (*1942*) call this phenomenon "mediated generalization," and they have drawn a rather complicated diagram—which is probably much too simple—to illustrate some relationships of this sort.

The verbal relationships diagramed in Figure 5.1 are logical and familiar to introspection. In order to obtain objective evidence for the reality of such relationships, Foley and Cofer (*1943*) set up a learning experiment in which recall of a list of words could be improved by

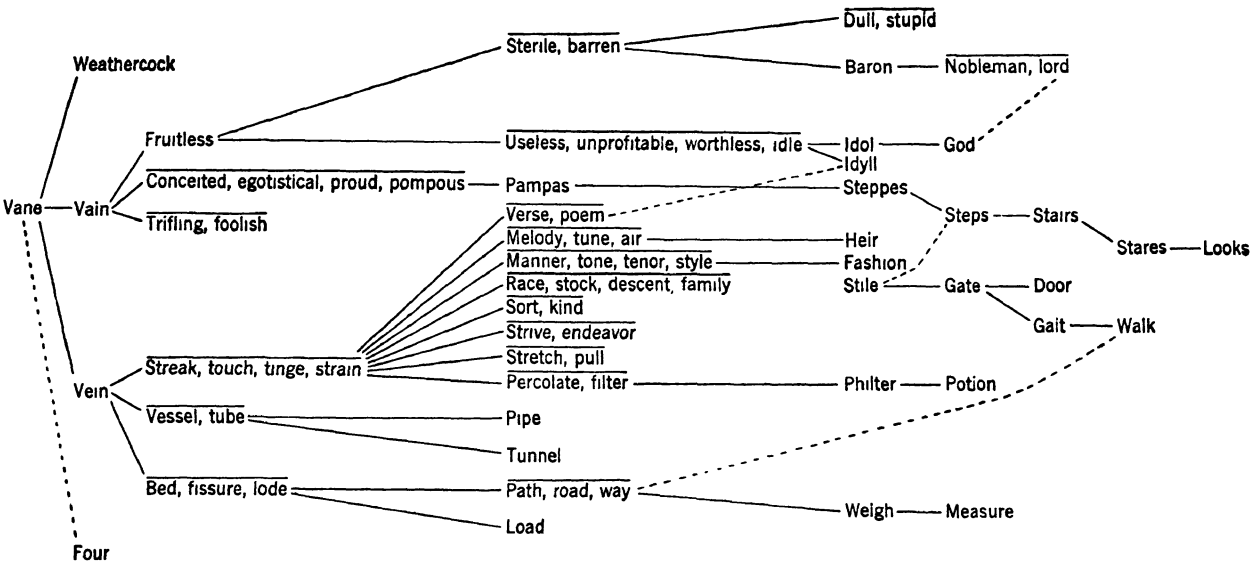


FIGURE 5.1. Formal representation of certain possible phonetic and semantic (mediated) generalization gradients. From Cofer and Foley (*1942*).

mediated generalization. They presented a group of college students with a list of ten short words, e.g., *sent,* to memorize. The subjects were then shown a second or reinforcement list of homophones of these words, e.g., *cent,* and then tested for recall of the original list. These subjects did better in recalling the original list than a matched control group that had the same original list followed by a reinforcement list of unrelated words. The same improvement was produced in another group of 25 subjects by a reinforcement list of different homophones, e.g., *scent,* and in another group by a list of synonyms, e.g., *dispatched.* Evidently the reinforcement took place indirectly, by mediated generalization or transfer along dimensions or gradients of homophonymity and synonymity.

Foley and Cofer also used a fourth list of words that were second-order synonyms. That is, for the original word *sent* a synonym is *dispatched,* and a synonym for *dispatched* (but not for *sent*) is *killed.* Exposure to this list of second-order synonyms also improved recall of the original list. It is hard to see any way that this second-order effect could occur except by mediated generalization or transfer along one gradient, than another.

Cofer, Janis, and Rowell (*1943*) have demonstrated similar first-order effects with antonyms.

To show how these gradients are built up Foley and Mathews (*1943*) used a list of ten English words and reinforcement lists of Spanish equivalents. Two groups of 40 students each recalled the English words equally well, but then the experimental group began a college course in Spanish. After several weeks the experiment was repeated, and now the experimental group was aided by the reinforcement lists of Spanish equivalents. The control group was not aided to any significant extent by reinforcement with the unfamiliar Spanish equivalents. Evidently the experimental group learned certain Spanish-English associations in the Spanish class which mediated reinforcement, in an indirect way, of the list of English words.

On the basis of these several experiments we can improve our concept of the supply of words. A person's storehouse of words may be pictured as arranged or sorted into gradients or clusters somewhat as merchandise is stacked in the stockroom of a department store or books are classified in a library. The analogy is misleading, however, because the arrangement of words is not spatial. The relationships between words, built up by previous experience, may be highly abstract

conceptual relations, speech-sound relations, Spanish-English associations, or even quite fortuitous relations.

Not only are the words in a person's vocabulary related by many kinds of relationships, but many words have multiple meanings and are therefore related to other words in different ways. It is a multidimensional structure of relationships. A librarian would say that there are many cross references. It is possible that these relationships could be disclosed by some variation of the sorting test, but apparently no test of this kind has been attempted. Certainly a multidimensional structure of semantic relations could not easily be erected in three-dimensional space.

In any event, when anyone approaches a verbal problem, he brings with him a supply of words and a structure of relationships among the words, all of which are consequences of previous learning. And some are more available than others for a specific problem because of previous learning of specific relationships.

INFORMATION: THE SUPPLY OF FACTS

Without doubt the supply of facts that anyone brings to a problem is an important factor in the outcome. Isolated facts are similar to isolated words in many respects and most of the previous discussion of words applies to the supply of facts. Methods for testing knowledge of facts in a limited area of learning are well known. Indeed it is common practice in educational psychology to test mastery of pertinent facts independently of ability to apply these facts to the solution of a problem. Typically, isolated facts are acquired by memorizing, incidental learning, and other simple forms of learning with which this book is not particularly concerned. But the importance of a supply of pertinent facts must never be overlooked.

CONCEPTS, IDEAS, GENERALIZATIONS: THE SUPPLY OF ABSTRACTIONS

Unfortunately the psychology of thought and judgment has been confused with the acquisition of such abstract material as concepts and principles. It is true that mastery of a difficult concept, either an artificial concept like "thun" or an authentic concept like "entropy," may require thought, and that concepts for this reason may furnish suitable material for experiments on thinking. Some of these experiments will be illustrated in a later chapter on productive thinking (Chapter 8).

But some concepts, like "sister" and "water," are acquired quite casually and thoughtlessly by a relatively simple form of learning of no special significance for thinking. And some thought problems, like jigsaw puzzles, are problems of spatial relationships having nothing special to do with the formation of abstract concepts. The important relationship for present purposes is that, when thinking about the problems of today, one makes use of all kinds of previously acquired materials, including supplies of sensori-motor coördination, facts, concepts, and principles. Therefore the kinds of abstract material available for thought should be briefly described.

CONCEPTS. The word "idea" has been used in psychology for many years in many ways. One usage is synonymous with thought, when thought means mental content, as in such phrases as "association of ideas" and "flight of ideas." The adjective "ideational" as used today often has the meaning of abstract or conceptual, as in "ideational learning" and "ideational fluency."

The dictionaries give several meanings for concept, but in psychology the usual reference is to the general concept—or the general idea. We say that someone has acquired the standard concept of *dog* when he can define dog adequately or discriminate dogs from a variety of objects that are not dogs. Many experiments have shown that the two methods do not always agree. Many people can discriminate numbers from non-numbers, but cannot define a number; many children have learned that "a verb is a word that asserts," but cannot sort verbs from nouns.

Socially accepted concepts are always designated by a communicable symbol, such as a word, though anyone may have a private concept that he will not or cannot communicate to others. Since many of the words included in a vocabulary test are names of general concepts, it is obvious that the score on such a test is related to the supply of concepts available to the subject, especially so if the words are abstract and the emphasis is on precision.

The tremendous utility of concepts to anyone is that they divide things up. They organize the world of uncountable objects, events, and ideas into a relatively small number of categories. If someone knows what is a *dax* and what is not, he can, with a trifle more effort, classify anything as *dax* or not *dax*. If he can use one thing that is a *dax* to solve a problem, he may be able to use another thing that is a *dax* in the same way. When he has the world divided into apples, binomials,

circuses, and so on to xylophones, yawns, and zanies, new things are no longer altogether new. He puts new things into familiar categories and transfers to them the effects of previous learning.

Simple concepts are organized into concepts of a higher level of abstraction, e.g., animal, vegetable, and mineral. The result, if the conceptualization process goes all the way, is a hierarchy or system of concepts, the best illustration of which would be a taxonomic system. Anyone who organizes the world he does business with into a consistent conceptual system has reduced tremendously the occasions for subsequent problem solving. If, in addition, his concepts are the standard ones of the culture, he makes available to himself a great range of information stored in the culture and increases the possibility of profiting by communication with other people.

GENERALIZATIONS. As an example of a generalization or principle we may take the familiar aphorism "What goes up must come down." Like other generalizations, it refers to two concepts, going up and coming down, and states a relation between them. Others, such as "the more the merrier," state a relation between variations along one scale and another, in this case a physical scale, number, and a psychological scale, merriment. Some generalizations are stated loosely, e.g., "Practice makes perfect," and some are stated precisely, e.g., $s = \frac{1}{2}gt^2$. Some are accurate and some are inaccurate. But people learn them and use them for better or worse in solving problems.

To test for the mastery of a generalization, we may ask that it be stated or that it be used. In general the two kinds of test yield the same results, but, as in the case of concepts, there are exceptions. Many students memorize a law without being able to apply it to new data. And, on the other side, when children throw a ball up, their behavior shows that they expect it to come down, but they may not be able to state the relation between the two events explicitly.

The solution of problems is influenced also by the thinker's supply of rules of action, which might be called procedural generalizations. Certainly the contestants in many games and in legal disputes bring to the contest previously acquired generalizations of the form "When this event occurs, you must do so and so." "Under these circumstances such and such an act is prohibited." Some of these procedures are very simple, like the operation indicated by $+$; and others may be rather involved, like the sequence of operations indicated by *do si do* and by

$\frac{(X-M)^2}{N}$. Here as elsewhere it turns out that people may be able to follow a rule of action without being able to state it. The reverse is also true. Especially in the area of moral procedures is the discrepancy between knowing what to do and doing it most apparent.

ATTRIBUTES, DIMENSIONS, SCALES. Another sort of abstraction of special importance for thought and judgment is the scale. Scales used for perceived objects have been discussed in Chapter 4, but the thinker approaching a new problem also brings with him many scales of abstract qualities conceptualized from previous experience. If we hand any ordinary adult a nail, a spike, and a tack, and ask him to arrange them in order of size, he will do it readily, thus giving evidence that the dimension of size is familiar to him. Furthermore, if we merely give him the names, nail, spike, and tack, he will be able to arrange them in order of size. If we name five books and ask him to rank them in order of merit, he will probably be able to do it—to his own satisfaction at least. When a foreman is asked to rate his men on the basis of their overall worth to the employer, he makes use of a scale of worth abstracted from specific past instances of behavior. The past instances of behavior differed from each other in many ways, but variations in worth to the employer can be abstracted from the complex and ordered on a scale. It is almost impossible to overemphasize the power of intelligent human beings to isolate an attribute or dimension and use it in thought and judgment. Like concepts and generalizations, dimensions may be acquired from others or discovered personally as the answer to a problem.

LIKES AND DISLIKES: THE SUPPLY OF AFFECTIVE MATERIAL

For the sake of completeness it is well to remember that everyone enters a problem situation with a repertoire of likes and dislikes. As problem solving proceeds, one acquires likes and dislikes for anything that appears to facilitate or retard progress toward the goal; but most people like and dislike many things from the outset, before any instrumental relation to the present problem is apparent.

These affective tendencies may be quite specific, like a preference for using a certain tool or a distaste for extracting square root. Some are more general, like a preference for a verbal or numerical solution—or a fondness for motor activity or intellectual contemplation. And preferences of a very abstract character may occasionally be observed, like

a mathematician's yen for an elegant solution or a diplomat's delight in solving international problems by balancing opposing powers. Such likes and dislikes will influence the solution of problems, especially difficult problems, and must be considered in any broad treatment of thinking. Considerable evidence on the effect of such affective tendencies on judgment is available, and some will be presented in a later chapter.

These likes and dislikes are, of course, the results of previous learning. They may be acquired by simple conditioning, by imitation and other social processes, and in some cases by more intellectual operations. Some likes and dislikes are quite superficial and may be disclosed by a questionnaire technique. Others are not readily admitted but must be inferred from behavior.

THE RAW MATERIAL OF THOUGHT

Facts, concepts, principles, and the like are often called the tools of thought, but the metaphor is a poor one. They are better called the stuff or raw material which the operations of thought fabricate into a judgment or the solution of a problem. These raw materials of thought are structured and some items are more available for use than others. Furthermore, recall is an active process. The thinker does not dip blindly into his supply. The dynamics of thought have a strong selective, sometimes distorting and editing, influence on what is remembered and used on a later occasion.

INDIVIDUAL DIFFERENCES IN LEARNING

If a hundred people attack a given problem, they will vary greatly in the success of their efforts. This variation is due in part to variations in present problem-solving abilities and in part to variations in supplies of sensori-motor coördinations, perceptual organizations, facts, and generalizations that they bring to the problem. And the variations in supplies are due to variations in learning abilities as well as opportunity for practice, motivation, recency of review, and the like.

Unfortunately standard discussions of learning have little to do with individual differences, although one of the most convincing reasons for studying the learning process is that it does make people different. The remainder of this chapter, therefore, reviews present knowledge of individual differences in learning abilities and achievements. Of par-

ticular interest is a question that came up earlier in this chapter, namely, the question of the relation between simple forms of learning and higher forms of learning or problem solving.

In general there are two kinds of scores that are important for the analysis of individual differences in learning and the relation between learning and thinking: scores obtained at the end of a period of standard opportunity to acquire new material, and scores based on uncontrolled past learning.

INDIVIDUAL DIFFERENCES IN AMOUNT LEARNED UNDER STANDARD CONDITIONS

Many experimenters have given their subjects an opportunity to learn two or more kinds of material and have correlated scores representing their achievements at the end of the practice period. Some of the published correlations between learning scores of this kind have been reviewed by Garrett and Schneck (*1933*), Hall (*1936*), and McGeoch and Irion (*1951*).

A simple memory factor, prominent in the memorizing of word pairs, names and faces, and the like, has been identified in large-scale factorial studies (Thurstone, *1938;* Thurstone and Thurstone, *1941;* Guilford, *1947*). It is one of the most independent of the factors extracted from the Primary Mental Abilities battery of tests, having a low correlation with the general factor.

When the learning required is of a simple arbitrary type, as in associating names and faces or pairs of words, correlations with tests of general intelligence are low. When the items are associated according to a code or pattern, terms like rational learning and mental maze are used, and correlations with general intelligence are higher (Garrett and Schneck, *1933*). Such code-learning tests are, in fact, included in several tests of general intelligence.

The difference between arbitrary and meaningful learning shows up also in the English, Welborn, and Killian (*1934*) studies of verbatim and substance memorization (see page 46). Scores for the two kinds of learning have low correlations, in the .40's, although the tests are fairly reliable and different tests of verbatim learning correlate in the .70's, as do different tests of substance learning.

The children in the Welch and Long (*1940*) experiment on levels of abstraction, described on page 236, were also given tests of number discrimination, matching, and memory span, five tests in all. Scores

on these tests correlated well, between .64 and .79, with success on the hierarchy tests. With age held constant, the correlations range from .42 to .66. Evidently the ability to learn the higher-order concepts is not a very special ability.

The term rational learning is used, not only for the mental maze and code learning, but also for the learning of the patterns of a bolthead maze. Roslow (*1936*) had 67 college students learn ten different patterns and obtained intercorrelations between scores for different patterns of .20 up to .69. Total scores on the first five correlated .67 with the second five given three months later. Each group of five problems correlated quite well with general intelligence, with the Minnesota Paper Form Board test, and with a syllogism test. Though the patterns were different in appearance, the results point to an important general factor having much in common with the general intelligence of conventional printed tests.

In summary the published correlations between learning achievements at the end of a standard practice period give little evidence for a general learning factor related to general intelligence. If such a factor operates throughout a variety of learning tests, it must be quite small and relatively unimportant. Much more important are group learning factors, such as one common to the learning of stylus mazes; another that shows up on cancellation tests; one of substitution learning; one of simple associative learning with only one exposure, commonly called the memory factor; and others of a more abstract or rational nature. In general those learning tests that call for mastery of abstract material correlate higher with tests of general intelligence.

Except for the memory factor these group factors are not well defined. The value of including tests of familiar reference factors has only recently been appreciated. If tests of familiar group factors, such as the verbal, numerical, spatial, and reasoning factors, had been included, the picture would be clearer. It may turn out that these group learning factors—except for the simple memory factor—have much in common with well-known factors that can be measured by conventional printed tests.

Learning and Problem Solving. We are now prepared to take up the question of the relation between learning abilities and problem-solving abilities. This discussion supplements the experimental analysis of this question reviewed earlier in this chapter.

Maier's experimental separation of learning and reasoning was

described earlier in this chapter. Campbell (*1935*) has studied this separation by an analysis of individual differences. He did not obtain a positive correlation between scores for reasoning and scores for learning either of two mazes, although a correlation of .55 between the two learning scores indicates that these scores have some reliability. Unfortunately the reliability of the reasoning scores is unknown, hence a negative conclusion is not very convincing.

If we take a broad view of learning, as recommended earlier, problem solving or reasoning is one kind of learning, so the question could be rephrased as a question of the relation between simpler and higher learning abilities. It has already been noted that a rational or abstract learning factor may be differentiated from simpler learning factors like memory ability. When the question of learning and problem-solving abilities is raised in connection with human subjects, psychologists usually contrast simple kinds of learning like memorizing and verbatim learning, in which practice is an important variable, to complex activities like working arithmetic problems, making inferences, and solving puzzles.

One difficulty in the investigation of this question is that of getting pure tests. Tests that look like tests of simple learning may involve some problem solving, and much problem solving makes use of past learning. If problem solving is the complex process it is supposed to be, a pure test of it may be impossible. It is true that ordinary tests of general intelligence consist largely of minor problems to be solved. For this reason they have been taken as measures of some sort of higher mental ability to be correlated with simpler abilities. These correlations have some pertinence to the question of the moment, but they cannot be clearly interpreted until the other components of the intelligence test, e.g., vocabulary, digit span, have been eliminated. As it is, our best evidence on the relation between individual differences in learning and problem solving would seem to come from correlations between scores on two tests that involve the same kind of material, e.g., verbal or numerical, but require different operations on this material.

Educational psychologists have occasionally compared scores on two examinations on the same subject matter, one testing simple learning and the other complex learning or problem solving. Billings (*1934*) constructed rather difficult tests of problem solving in eight fields, such as arithmetic, physics, and sociology; but before the subjects, 146 college students, took these tests, they studied the facts and principles

on wh:h the problems were based and took true-false tests on this materi . The eight correlations between the information tests and the problem-solving tests ranged from .35 to .59, with a median of .46. The problem-solving tests were shown to be fairly reliable, but there is some question about the reliability of the information tests. If corrected for attenuation, the median correlation would probably be in the .50's. Scores on the eight information tests were only slightly related to intelligence (Army Alpha), with a median correlation of .37, while the corresponding median for the problem-solving scores was .54. There was a large sex difference in favor of the men on the problem-solving tests, though the sex differences on information and intelligence were small.

Bedell's (*1934*) investigation furnishes a rather clear-cut picture of the relation between learning and problem solving within a small area, general science. The test of learning was not one of continued practice but involved reading a paragraph about general science and then, without being able to see the paragraph, answering factual questions about the contents of the paragraph. The problem-solving test required making inferences from the paragraph. Both kinds of item were in multiple-choice form. Obviously any difference in the scores on these two tests must depend on a difference in intellectual processes rather than on the form of the items or the kind of material. The whole test, 30 items of each kind, along with the Terman group intelligence test, was given to 324 students in the eighth and ninth grades taking a general science course.

Here again an interesting sex difference appears in the results. The girls were slightly and not significantly better than the boys on the intelligence and recall tests, but the boys were significantly better than the girls on the inference test. The correlation between recall and inference was .65, or, if corrected for attenuation, .77. As one would expect, the intelligence test correlates higher with inference than with recall when both are corrected for attenuation.

In summary, then, we must conclude that simple learning and higher learning or problem solving involving the same kind of material have much in common in respect to the production of individual differences. The correlation coefficients are in the middle range. The two kinds of abilities could not, however, be considered the same. This conclusion agrees with the conclusions from experimental studies reviewed earlier in this chapter.

INDIVIDUAL DIFFERENCES IN PRESENT ACHIEVEMENT BASED ON UNCONTROLLED PAST LEARNING

Another way of studying individual differences in learning is simply to inventory present knowledge and skills. What anyone can do today depends upon learning abilities, past opportunities to learn, motivation, recency of review, interferences of various kinds, and other uncontrolled factors. When a hundred people began to think about anything, individual differences in the outcome will depend largely on individual differences in present supply of learned accomplishments, resulting from much uncontrolled past learning.

This relationship can be stated a priori in a negative way. If the solution of a problem requires certain factual knowledge, those thinkers who have not had the requisite ability to acquire, retain, and recall such knowledge, *and* the opportunity, and the motivation at some time in the past, simply will not solve the problem. The others may or may not, depending upon problem-solving abilities functioning at the moment. We can add to this a priori negative relationship some information on various supplies and accomplishments that different people may be expected to bring to various problems.

It is not feasible to sample the sensori-motor habits of a hundred subjects, but if it were done the correlations would probably be low. We might expect to find occupational similarities and differences however. People accustomed to working with small tools, for example, might have a different supply of skills available for the solution of a mechanical problem than those accustomed to heavy tools, or those unaccustomed to tools. Some people may habitually turn the puzzling object around, or begin at the upper left corner, or poke at it with a pencil. Children are more likely to use their teeth than adults. And some children, because of family training, are unusually reluctant to touch things, especially things that are dirty or sharp. Special problems could be devised in such a way that mastery of a particular sensori-motor coördination would be critical for solution.

There is little evidence on individual differences in perceptual achievements pertinent to the solution of problems, but here also one would expect occupational differences. A radiologist examining an x-ray of his patient's chest, an experienced mechanic listening to a defective motor, a cook opening the oven to judge the status of the apple pie, all bring patterns of sights, sounds, and smells from the past which help or hinder diagnosis of present problems. Scores on tests of

mechanical ability depend in part on experience in perceiving tools, relations between pulleys and belts, and signs of stresses.

Considerable information is available on individual differences in the supply of words, both general vocabulary and vocabularies in special fields, some of which has been included earlier in this chapter. Two tests of general vocabulary of the same sort but of different words will usually correlate well with each other, hence it is defensible to speak of individual differences in the supply of words in general. There is some evidence for a general factor (Johnson, *1941*). One argument for the importance of past verbal achievements in problem solving is the usefulness of tests of general vocabulary and especially of technical vocabularies in the prediction of job success.

What is true of words is probably also true of factual information. The fact that it is possible to construct reliable tests of general information, as on the Bellevue-Wechsler Scale of Intelligence, indicates that there is some generality of individual differences over a broad sample of the population in respect to supply of information. It is likely, however, that the general factor is quite small and group factors prominent. Tests of technical information or trade knowledge are useful predictors of success in solving problems on the job (Stead, Shartle, *et al., 1940*).

Since tests of vocabulary and information predict success on the job moderately well, one is tempted to take this relationship as proof of the importance of past learning for the solution of present problems. But there is another possibility. A supply of words and facts about aviation indicates an interest in this field—in fact, information tests have been used as interest tests—and the continuing interest, as much as the supply of information, may be responsible for the job success.

Individual differences in supply of concepts are tested to some extent by vocabulary tests and by tests of school achievement. In addition, educational and social psychologists have studied the mastery of certain concepts and principles of special interest. Since the pioneer work of Piaget children's concepts of causal relations and the like have attracted considerable empirical attention. Much of this work on concept formation has been recently brought up to date by Vinacke (*1951*).

Past Learning and Present Problem Solving. The relationship between problem solving and achievement based on uncontrolled past learning has often been studied in two areas of learning, verbal and numerical. When tests that require reading and interpreting a difficult paragraph, called reading comprehension or inference, are compared

with vocabulary tests, the correlations are moderately high, usually in the .50's. When tests of arithmetic problem solving or reasoning are compared with tests of numerical computation, the correlations are also moderately high, usually in the .40's.

Within any limited area of knowledge it is safe to conclude that the successful problem solvers will have larger supplies of previously learned facts and principles than the unsuccessful. But the relationship is not close. Problem solving depends on previous learning, but there is more to it than that. The next few chapters will be directed specifically toward those higher intellectual activities—called problem solving, reasoning, complex learning, productive thinking, evaluating, judging, and what not—that make use of previous learning but go beyond it.

REFERENCES

Bedell, R. C., 1934. *The relationship between the ability to recall and the ability to infer in specific learning situations,* Kirksville, Mo., Bulletin of the Northeast Missouri State Teachers College, 34, No. 9.

Billings, M. L., 1934. Problem-solving in different fields of behavior, *Am. J. Psychol., 46*:259–292.

Campbell, A. A., 1935. Community of function in the performance of rats in alley mazes and the Maier reasoning apparatus, *J. Comp. Psychol., 19*:69–76.

Cofer, C. N., and Foley, J. P., Jr., 1942. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena, *Psychol. Rev., 49*:513–540.

Cofer, C. N., Janis, M. G., and Rowell, N. M., 1943. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients, *J. Exper. Psychol., 32*:266–269.

English, H. B., Welborn, E. L., and Killian, C. D., 1934. Studies in substance memorization, *J. Gen. Psychol., 11*:233–260.

Foley, J. P., Jr., and Cofer, C. N., 1943. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients, *J. Exper. Psychol., 32*:168–175.

Foley, J. P., Jr., and Mathews, M. A., 1943. Mediated generalization and the interpretation of verbal behavior: IV. Experimental study of the development of inter-linguistic synonym gradients, *J. Exper. Psychol., 33*:188–200.

Garrett, H. E., and Schneck, M. R., 1933. *Psychological tests, methods, and results,* New York, Harper & Brothers.

Goodenough, F. L., 1942. The use of free association in the objective measurement of personality, in *Studies in personality,* ed. Q. McNemar and M. A. Merrill, New York, McGraw-Hill Book Co.

Guilford, J. P., 1947. *Printed classification tests,* Army Air Forces Aviation Psychology Program Research Reports, Report No. 5, Washington, D.C., Government Printing Office.

Hall, C. S., 1936. Intercorrelations of measures of human learning, *Psychol. Rev., 43:*179–196.

Hilgard, E. R., 1948. *Theories of learning,* New York, Appleton-Century-Crofts.

Hull, C. L., 1935. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution, *Psychol. Rev., 42:*219–245.

Johnson, D. M., 1941. Confidence and achievement in eight branches of knowledge, *J. Educ. Psychol., 32:*23–36.

Katona, G., 1940. *Organizing and memorizing,* New York, Columbia University Press.

McGeoch, J. A., 1942. *The psychology of human learning,* New York, Longmans, Green & Co.

McGeoch, J. A., and Irion, A. L., 1951. *The psychology of human learning,* New York, Longmans, Green & Co.

Maier, N. R. F., 1929. *Reasoning in white rats,* Comp. Psychol. Monogr., No. 6.

Maier, N. R. F., 1931. Reasoning and learning, *Psychol. Rev., 38:*332–346.

Maier, N. R. F., 1940. The behavior mechanisms concerned with problem solving, *Psychol. Rev., 47:*43–58.

Maier, N. R. F., 1945. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning, *J. Exper. Psychol., 35:*349–360.

Maier, N. R. F., and Schneirla, T. C., 1935. *Principles of animal psychology,* New York, McGraw-Hill Book Co.

Melton, A. W., 1941. Learning, in *Encyclopedia of educational research,* ed. W. S. Monroe, 667–686, New York, The Macmillan Co.

Razran, G., 1939. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning), *Science, 90:*89–90.

Razran, G., 1949a. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli, *J. Exper. Psychol., 39:*642–652.

Razran, G., 1949b. Some psychological factors in the generalization of salivary conditioning to verbal stimuli, *Am. J. Psychol., 62:*247–256.

Riess, B. F., 1940. Semantic conditioning involving the GSR, *J. Exper. Psychol., 26:*238–240.

Roslow, S., 1936. A statistical analysis of rational learning problems, *J. Genet. Psychol., 48:*441–467.

Seward, J. P., 1949. An experimental analysis of latent learning, *J. Exper. Psychol., 39:*177–186.

Stead, W. H., Shartle, C. L., *et al.,* 1940. *Occupational counseling techniques,* New York, American Book Co.

Swenson, E. J., Anderson, G. L., and Stacey, C. L., 1949. *Learning theory in school situations,* University of Minnesota Studies in Education, No. 2, Minneapolis, University of Minnesota Press.

Szekely, L., 1950a. Knowledge and thinking, *Acta Psychol., 7:*1–24.

Szekely, L., 1950b. Productive processes in learning and thinking, *Acta Psychol., 7:*388–407.

Terman, L. M., and Miles, C. C., 1936. *Sex and personality: studies in masculinity and femininity,* New York, McGraw-Hill Book Co.

Thurstone, L. L., 1938. *Primary Mental Abilities,* Psychometric Monog., No. 1, Chicago, University of Chicago Press.

Thurstone, L. L., and Thurstone, T. G., 1941. *Factorial studies of intelligence,* Psychometric Monog., No. 2, Chicago, University of Chicago Press.

Underwood, B. J., 1949. *Experimental psychology,* New York, Appleton-Century-Crofts.

Vinacke, W. E., 1951. The investigation of concept formation, *Psychol. Bull., 48:*1–31.

Welch, L., 1947. The transition from simple to complex forms of learning, *J. Genet. Psychol., 71:*223–251.

Welch, L., and Long, L., 1940. The higher structural phases of concept formation of children, *J. Psychol., 9:*59–95.

Chapter 6

Preparation for Thought and Judgment

A question well put is half answered.

EVERYTHING that precedes and influences thought may, of course, be called preparation for thought. The contribution of past learning, as described in the previous chapter, is preparation for thought when preparation is considered in a static sense. Preparation in a dynamic sense is a process, the process of getting ready or adopting a preparatory set, based on present conditions as well as past learning, which controls the subsequent production of pertinent responses. Several kinds of evidence, from experiments on association and problem solving as well as studies of creative thinking, permit the identification of a period of preparation preceding the productive process. The development and operation of the preparatory set was discussed in a general way in Chapter 3. It is time now to attempt to describe how people prepare for various intellectual enterprises. The psychology of the preparation for thought is largely an elaboration and application of the psychology of set.

THE DISCRIMINATION EXPERIMENT

Typically the subject is prepared for the discrimination experiment, as for the reaction-time experiment, by instructions from the experimenter. He is told, for example, that at a certain time two lines will be projected on a screen; that he should be ready to observe them; that he should look at their lengths, not their widths or colors; and that he should judge the left one in reference to the right one as

"longer" or "shorter." From instructions like these, usually quite explicit, the subject builds up a set, that is, a readiness to perceive certain stimulus patterns and to make certain responses.

In addition to such explicit instructions the subject is influenced by his general orientation to the experimental situation. His perception is influenced by the screen and its background, stray lights and other irregularities, and possibly by recall of similar perceptual assignments in the past. His response is influenced by his knowledge of the meaning of "longer" and "shorter," his interest in the experiment, and his plans for the next hour. Nevertheless, if the experimental instructions are adequate, the subject is prepared to abstract from the complex experimental situation only those stimulus aspects with which the experiment is concerned and to select from his supply of responses only those requested.

Under ideal conditions the set may control the subject's activities so well that the variables of thought and judgment are held constant and the responses may then be directly related to the operation of the perceptual mechanisms. This is the aim of psychophysical experimentation, and a precise methodology has been developed over a period of many years in order to hold constant these extraneous variables of judgment. On the other hand the experiment may be redesigned so as to emphasize the effect of the set on the subject's responses and thus to illuminate the judgment process. Later chapters will review the evidence from experiments of this type.

THE ASSOCIATION EXPERIMENT

Words are important to thought, carrying meaning in small discrete packages, easy to record and count, and relatively easy to classify. The association experiment in its many forms, dealing chiefly in single words, yields considerable information on preparation for a rather simple form of thought, bridging the gap between the simplest reactions, as in reaction-time and discrimination experiments, and the complex enterprises of problem solving and creative thought.

FREE ASSOCIATION

Typically, in the free association experiment the subject is given his set by instructions to say the first word he thinks of; then he is given the stimulus word, which might be *rain*. The response word, after an association time of a second or two, might be *snow*. Here is an experi-

ment concerned with relatively significant matters; yet, like the reaction-time and discrimination experiments, it is one in which the preparation period, before the presentation of the stimulus word, can be definitely marked off from the main period. Nevertheless, it is difficult to discover what goes on during the preparation period, aside from rather nonspecific data on muscle tension, referred to in Chapter 3. Our information comes chiefly from *ex post facto* analysis of relations between the response word and the stimulus word and from introspective reports. Most of the research on word association was done before 1930 and is available in Woodworth (*1938*).

Classifications of Associations. Many attempts have been made to group the associations into a few large classes based on the relationship of the response word to the stimulus word. The relationship between any one response word and its stimulus word may be due to some peculiarity of the stimulus word, but when many stimulus words are used such peculiarities are likely to be balanced out of the totals. Consequently, if the classification procedure is a good one, any trends that appear can be taken as indications of what the subject is prepared to do at the moment the stimulus word is presented to him.

Aside from the fact that people who give many coördinates are likely to give many contrasts, there has been no evidence for the reliability of such classifications until recently. A reliability study by Karwoski and Berthold (*1945*) makes use of the following classification of relationships:

Essential similarity: large-big, rough-rugged.
General identification: cabbage-vegetable, hand-arm.
Specific identification: ocean-Pacific, friend-Tom.
Contingent identification: egg-breakfast, music-roommate.
Essential opposition: hot-cold, fill-empty.
Contingent opposition: foot-hand, house-barn.

There are several ways of checking the reliability of such a classification. Karwoski and Berthold had the responses classified by four analysts and calculated the interanalyst agreement between one pair of analysts and the other pair. When the responses of 100 subjects to 20 words were classified, there was disagreement between the two pairs of analysts to the extent of 18 percent of the responses in one study and 20 percent in another study. Correlational analysis shows that the classes are not independent. The percentages of 5811 responses of these college students which were classified in the above categories of rela-

tionship were, in the same order as the table above, 15, 13, 17, 30, 19, and 6. Apparently the subjects were most frequently set for contingent identification of the stimulus words and least frequently for contingent opposition.

Three years later Karwoski and Schachter (*1948*) used a slightly different classification system. The frequency of each and association times are shown in the table below:

	General identification	*Specific identification*	*Contingent relation*	*Essential opposition*
Percent frequency	31	3	23	43
Mean time in seconds	1.34	1.52	1.44	1.10

Comparison of the frequencies in the two experiments shows that the number of responses placed in any one category depends on the other categories that are available. When the later, fourfold set of categories is used, essential opposition is the most frequent type of relationship used by college students. The table above adds some support to Marbe's law (see page 179) in that those responses given most frequently are given fastest. The time differences are statistically significant. The finding of significant time differences backs up the validity of the classification system because if the classication did not have some psychological meaning, time differences could be only chance differences.

These studies have all been done with words as stimuli and words as responses. Some evidence is available, however, from Karwoski, Gramlich, and Arnott (*1944*), using objects, pictures of these objects, and words naming these objects as stimuli, and requiring words as responses to all three types of stimulus. In respect to frequency of responses in the various classifications of relationship, the three types of association are similar, with minor exceptions. In respect to speed, responses to the words were fastest on the average and responses to pictures slowest, but those stimulus items which evoked a fast response when presented one way also evoked fast reactions when presented the other ways, with other subjects, as indicated by correlation coefficients of approximately .50. It would not be correct to maintain that the results obtained in word-association experiments are peculiar to the use of words as stimuli.

These three recent investigations do not carry us far in our study of

the preparatory set for thinking of words but they show that sophisticated critical analysis of word association is possible.

The set for word association changes with practice in a fairly regular way. The task becomes less of a challenge and the responses become faster and more superficial. In an experiment by F. L. Wells (1911), for example, the word "axle" evoked the response "hub" on first presentation but, when presented again after practice with a thousand words, the response was "grease." Likewise the response to "spread" changed from "distance" to "bed." Habituation to this kind of experimental situation is a change in the subject's preparation for response, established before the stimulus word is presented.

RESTRICTED ASSOCIATION

In free association the subject's preparation is relatively nonspecific and must be inferred *ex post facto* from relations between response words and stimulus words. In restricted or controlled association, by contrast, the preparation is adopted by the subject at the experimenter's request. The preparatory set can thus be experimentally varied—within limits, to be sure, of the subject's larger motivational pattern—and is in fact the principal independent variable of the controlled association experiment. Another difference between the two types of association experiment is that in the latter, when the instructions are specific, the responses can be right or wrong. This possibility of being wrong introduces changes in the subject's overall motivation within which the preparatory set operates, especially when the restriction is so rigid that only one response can be right.

The following are typical tasks:

Synonyms. Name a word that means the same as each of these words.

hard
level
coruscated

Species-genus. Name a word for the class in which each of these words belongs.

shovel
house
evolution

Add the next term.

Monday, Tuesday .....

2, 4, 6 .....
9 + 3 =
incisor, canine .....

From these illustrations it is apparent that the degree of restriction can be varied considerably. The more restricted the task, the easier it is to classify the response word as wrong and the more the task takes on the character of a problem. Such tasks may in fact be minor problems in themselves or parts of larger problems, such as working a crossword puzzle, studying a school assignment, writing a poem, and making out an income tax report. If the task is difficult, the subject is likely to take a critical attitude toward the words he has produced; a retroflex activity of judgment occurs. The subject turns round to verify the agreement between the solution and the requirements of the task. Thus we have in miniature a complete act of thought, preparation for which is well worth scientific investigation.

Certainly the remote preparation or education for the task should not be overlooked. A person who has not learned the meaning of "coruscated" is not well prepared to name a synonym. But the central problem of the psychology of thought is to describe how, when a subject is familiar with these words as well as many thousands of other words, he can prepare himself in such a way that the stimulus word is soon followed by an appropriate response word.

The objective facts are quite remarkable. Ask any ordinary literate adult to be ready to name opposites. Then give him the word "hard." Of all the words in his vocabulary, perhaps 100,000, he is likely to select one, probably "soft" or "easy," and announce it within two or three seconds. If the task is more difficult, he may take more time and his reply may not be precisely accurate, but in most cases the response word will be more or less appropriate. This selectivity of thought has intrigued psychologists for many years, and names such as determining tendency and *Einstellung* have been given to whatever it is that causes the responses to be relevant to the requirements of the task. The name of course is unimportant. The activity clearly conforms to the customary definition of set, a readiness to make a certain response to a certain stimulus. In the interest of economy of terminology, the term set is to be preferred.

The very fact that preparation by instructions does control response is almost the only objective fact available. To prove that time is required for preparation for word association, one could give the

preparatory instructions and the stimulus word simultaneously, and note the response time for comparison with response time when the stimulus word follows a preparation period. To prove that attention is required for preparation for word association, one could try the distraction procedure. That is, one could study the effect, in respect to lengthening the association time, of distraction during the foreperiod. Perhaps the foreperiod is more vulnerable to distraction than the main period or, in other words, perhaps a higher degree of attention is required when preparing for word association than when actually responding to a prepared-for stimulus word.

Such analytical distraction experiments apparently have not been performed, but Jersild's (*1927*) experiment on set and shift required alternation of simple association tasks, such as simple addition and naming opposites, with time of total performance as the dependent variable. One might consider these two tasks in rapid succession as mutually distracting, and in fact alternating usually is slower at the outset than executing each task separately.

Jersild's subjects, who were college students, soon became accustomed to the alternation, showing that the alternate sets, once organized, can be quickly put on and off as units or, in other words, that one can acquire a higher-order or compound set. There was interference, however, when the alternate tasks were similar, requiring alternate addition and subtraction of the same number.

Wells's prolonged experiment (see page 150) demonstrated that practice changes the set for free association, making the responses more superficial. Razran (*1949*) has deliberately had his subjects practice restricted association, e.g., naming contrasts and supraordinates, just before a conditioning experiment, and found that such practice significantly influenced the generalization of semantic conditioning (see page 129).

An experiment by Herrick (*1939*), though devised for another purpose, fits in at this point because it bridges the gap between the association experiment and the problem-solving experiment. He instructed his subjects as for the free association experiment. That is, they were to respond with a word, or in some cases with a number from 1 to 6, when they heard the stimulus word. In addition the experimenter would say "Right" or "Wrong" after each response without indicating the basis of evaluation. Thus the subjects had the problem of trying to puzzle out the kind of responses the experimenter wanted. In one

experiment, for example, the right response would be a habitual speech sequence, such as *darling-daughter* and *September-October*. In another the right response would be 1 if the stimulus word was a noun, 2 if it was a verb, and so on. Though the experiment has the form of free association, it could also be considered a problem-solving experiment in which success was manifested by discovery and adoption of the correct set for controlled association—a task-induced set. Most of the six subjects were able to do the easier problems; none got the hardest one. Their procedure usually was to try on various preparatory sets tentatively for a few trials—hence they could be called hypotheses—and test the results. We shall see in the next section on problem solving that the discovery of the appropriate set may be a problem in itself, after which production of the correct response is automatic.

How the Preparatory Set Operates. It is obvious that the concept of preparatory set for word association is a useful one which seems to designate some definite psychological function occurring before the stimulus word is presented. The question of the operation of this preparatory set in selecting some specific response from the large supply of available responses is the central question of the psychology of thought, comparable to the question of reinforcement in the psychology of learning.

The introspective comments of competent subjects during preparation for word association indicate that they are often aware of an incomplete pattern or blank ready to be completed. This schema is in part given to the subject by the experimenter's instructions and in part constructed by the subject when he hears the stimulus word. Having this schema "in mind," the subject searches his memory or supply of words for one that fits. Thus the schema may be pictured as a model or template which guides the productive process or flow of associations. Some subjects report visual images of search models in the form of diagrams of the relationship with a gap indicated. That people do use search models of this sort in thinking is an altogether reasonable assumption and, as Woodworth says, "this concept lays claim to a position of fundamental importance in the psychology of thinking."

The principal difficulty with this concept, of course, is that it developed from introspection, a notoriously unreliable source of evidence. Much may happen that is not reported. The psychology of thought is so improverished, however, that it cannot be unduly fastidious about its sources of nourishment. At the present time it

would be worthwhile to treat the schema or search model as a research possibility and try to find more objective evidence to support it —or refute it. The difficulty is not that the evidence is introspective, but that it is only introspective.

There are at least two possibilities for making the evidence on the search model more objective. One is a comparison of responses guided by clear and unclear search models. Instructions of different degrees of adequacy can be given to the subjects and they can be told to prepare to make the response. Before actually making the response, they might be told to rate the clarity of their search models as "clear" or "vague." If the responses are faster or better with the clear than the vague search models, the function of the search model is supported. If the subjects cannot rate the clarity of their search models, or if clear search models do not lead to better responses than unclear search models, doubt is cast on the concept of search model.

The reproduction method offers another possibility. Perhaps experiments could be arranged in which the subjects make drawings or schematic diagrams of their preparations in some way. The matching technique may be used if diagrams are not feasible. Any procedure by which subjective data could be compared with objective results has heuristic potentialities.

As the evidence stands now, the search model is as much a logical as a psychological concept. The subjects, well-educated thinkers, report what they know, consciously, must be done. What they do may or may not coincide with their picture of what should be done or of what people would logically be expected to do.

If the schema is accepted as a logical concept which probably corresponds in some ways with psychological operations, a broadened account of its role in word association may be attempted. From the whole environment, perceived as a whole or in parts, at present or in the past, the thinker abstracts certain aspects connected in some way, perhaps logically, perhaps accidentally, with the present task. The experimenter's explicit instructions are prominent aspects of the environment, to be sure, but inferences from the experimenter's behavior, generalizations from similar social situations, the effects of practice, and of course the subject's motivation for doing all this must not be overlooked. The thinker combines these influences according to his own abilities and biases into a schema. With this schema as a guide or

model he searches his memory for a word that fits. The search model, then, is the conscious aspect of the preparatory set.

In this account, or any other reasonable account of the role of the search model, it appears that the search model, which is conscious, guides associations, facilitating some responses and inhibiting others before they reach the threshold of consciousness. This seems paradoxical if we accept consciousness as our criterion. How can a person know that a word is suitable or unsuitable before he knows what the word is? It would seem that a separate act of judgment would be necessary; that words would have to be produced some way, perhaps at random, then tested one by one for appropriateness. Yet the introspective reports reveal nothing of the kind. Only words more or less relevant enter consciousness. Subjectively, the process is quite direct, yet reasonably accurate.

This paradox is not a paradox, however, if we ignore introspective data and decline to trust reports of the thinker's consciousness. We can trust the objective records of the thinker's responses and the instructions given to him, and we have some objective experimental evidence on conditions that influence the association time and the kind of response. We can make reasonable assumptions about a hypothetical searching or scanning process which will yield the results of association experiments without making any assumptions about conscious awareness of any phase of the process—except perhaps the final announcement of the response word. Those who like mechanical models will find it easy to picture a machine that will scan a supply of information punched on cards and select one or more which fit the pattern for which the machine is set.

LABORATORY AND CLASSROOM PROBLEMS

How many words can you make out of PSYCHOLOGY?
A man and his wife together are 98 years old. He is twice as old as she was when he was the age that she is now. How old is each?
Take off a man's vest without removing his coat.
If no *X*'s are *Y*'s and no *Y*'s are *Z*'s, can an *X* be a *Z*?

Much of our knowledge of thought comes from observation of human and animal subjects working the problems offered them in the classroom or the psychological laboratory. The problems may be ordinary arithmetic problems; verbal problems, such as anagrams and crossword puzzles; or spatial problems that require locomotion or

manipulation in spite of difficulties. With animals and children who do not understand verbal instructions, the usual procedure is to use an incentive or lure which motivates the subject and also defines the solution, thus establishing the preparatory set. When the subjects are human adults or older children, they are prepared for the problem by verbal instructions.

Human subjects solve problems under the pressure of many kinds of motivation. In the laboratory and classroom the motivation is usually social. They work because they want to be agreeable and satisfy the experimenter's or the teacher's expectations, or because they want to surpass other subjects or test out their own abilities. Nonsocial motivation must not be overlooked, however, as pointed out in Chapter 3. Some people do undertake a solitary struggle with a crossword puzzle and some do try to anticipate the ending of a mystery story—driven apparently by nothing more than idle curiosity or love of intellectual adventure. And some students deliberately pick courses that offer an intellectual challenge. If they are asked why they put forth so much effort, they often reply, "Just for the fun of it," or "I would like to see what it is all about," or "I want to find out what happens." It seems quite likely that such an interest in the material of the problem for its own sake complicates and occasionally dominates preparation for problem solving in the laboratory and classroom.

This matter of motivation may not be trivial. A subject who gets ready for an arithmetic problem just to please someone else will be ready also for activities other than problem solving that might please his taskmaster. Another subject who enjoys playing with arithmetical relationships may become so absorbed in arithmetical manipulations that he loses sight of the goal he was told to work toward. The one will be distracted by social factors while the other will be distracted by the charm and potentialities of his material.

Whatever the motivation may be, preparation for problem solving is more complex than for the simple forms of thought thus far discussed. There are several reasons for this. First, in experiments on reaction time, simple discrimination, and word association the subject gets set; then, when the stimulus is presented, he makes his response. In such a one-cycle episode the preparation period is clearly marked off from the response period by the presentation of the stimulus. But if a problem is at all difficult, the preparation is not so simple. The subject may say, "I don't even know how to begin." Or he may get

ready to make certain movements or to perform certain arithmetical manipulations; then, if not rewarded, he may revise his preparation and get ready to make other movements. We shall find that in multicycle problem solving the preparation, though it may be identified by the subject or another observer, cannot always be temporally localized in one initial period.

Second, the time span is longer. In the simpler forms of thought the subject is prepared, the stimulus comes, and he gives his response. The burst of thought is all over, for better or worse, in a few moments. The execution of the response resolves the set, releasing the thinker for routine activities. Problem solving, on the contrary, takes time, and during this time much can happen. The set, like any other characteristic of the organism, fluctuates from moment to moment. Thus it becomes necessary to consider temporal aspects of the set, its function in steering intellectual activities over a period of an hour or so, resisting or succumbing to distraction from within and without. This will take us into such topics as control of attention, boredom, satiation, and fatigue, which have been discussed in a general way in Chapter 3. Furthermore, almost any response resolves the set for word association, but the execution of a response or solution to a problem may be merely the beginning of a different, more critical intellectual operation: judgment of the solution.

Third, the preparation for problem solving is much less explicit than for simpler forms of thought. In word association the instructions prepare the subject quite definitely for the forthcoming stimulus word and the desired response word, but the instructions for an arithmetic problem are much less helpful: "Find a number which when multiplied by itself gives twice the first number." Most subjects will not produce numbers immediately but will study the problem first in an overall way. Or the instructions may be simple enough, e.g., "Separate the two rings," and the subject may proceed at once to overt responses, discovering only later how to specify the response sought.

According to our earlier definition of problem (see page 63), the subject's first response is not rewarded. If it is a complex problem, none of the responses produced by the operation of the first set is rewarded. We may say that the first set or preparation for the problem is unrewarded. Thus it is characteristic of complex laboratory and classroom problems that the preparation, specifying the stimulus patterns to look for and the responses to make, is a subsidiary problem in itself.

In solving these problems the thinker contributes much more to the preparatory set than in the simple forms of thought, instructions for which establish a set that permits immediate production of responses.

Is it profitable, in consideration of these three differences, to assume that some sort of preparatory set influences the solution of complex problems? Logically the set would seem to play the same role in complex problems as in simple problems. The responses produced during problem solving are certainly not random any more than word associations. No competent observer has claimed that his animal or human subjects in a problem situation were as likely to make one response as another. Something like a set makes some responses more probable than others. As in word association and reaction time experiments, the set is geared to the organism's overall motivation and is susceptible to distraction. It seems reasonable to suppose that picking certain aspects out of the problem situation, and integrating them into a pattern or schema which then defines and guides the production of responses, is a separate process from the production of responses. We shall see that the crucial part of some problems is just this preparation; that if one prepares in some ways production of the correct response is easy, almost automatic, and if one prepares in other ways production of the correct response is impossible. It follows also that as problems become complicated, the capacity of the organism for integrating initially disparate aspects of the problem situation becomes a limiting factor.

How, then, can the preparation process be identified in the solution of complex problems? If the above analysis is correct, the preparation may not be confined to a short period preceding the production of responses. Problem solving is cyclical, preparation and production often alternating. But it follows from the above analysis that under control of one preparatory set the responses produced would be similar. If a subject doing a geometrical problem make six responses of one kind, e.g., draws six triangles, then makes five responses of another kind, e.g., writes five formulas, we may assume that his set changed. It may be difficult at times for an observer to classify six responses as of one kind and five as of another, but we shall see that it is not impossible.

It becomes necessary, if the foregoing account is approximately correct, to study two complementary aspects of the preparation for complex problem solving: surveying the problem, and setting up the requirements for solution. If the problem is simple or if the organism engaged in it is simple, it may not be necessary to carry the analysis this

far. In the case of unicellular organisms, reacting tropistically, even the differentiation of perception and response is somewhat forced. But as the organisms and their problems become more complicated, their activities become more highly differentiated. This book is concerned with the most highly differentiated organism known, engaged in the most highly abstract, involved, and subtle intellectual operations on record. We must be willing to carry the analysis as far as necessary. The risk of undue fragmentation is balanced by the risk of superficiality.

SURVEY OF THE PROBLEM SITUATION

One could prepare for problem solving by exploring the problem situation or by studying the required solution, by examining what is given or what is required. Perhaps both are done at once, but the different sides of the preparation may be emphasized at different times, by different thinkers, by different problems, or by different modes of instruction. The distinction is clearest in the case of arithmetic problems, where what is given is often collected and reinterpreted or rehearsed before study of what is to be found. Teachers of arithmetic sometimes train their students to assemble the data of the problem in a systematic way, and examine the implications of the data before proceeding further.

When a laboratory animal is placed in an unfamiliar cage, maze, or problem box, he usually explores his surroundings, sniffing, pawing, or otherwise familiarizing himself with the general layout of his theater of operations, locating prominent objects, barriers, holes, light and dark places, often manipulating things that are movable. Objects embedded in an inclusive pattern are overlooked; things that move and things that are brightly lighted evoke exploratory activity. This is perhaps the most primitive action tendency of all. It takes place quickly, and it seems so natural and nonspecific, operating about the same way whatever the animal's motivation, that many experimenters overlook its importance and leave this preliminary survey out of their records of problem-solving activities. Nevertheless the animal very quickly registers the illumination, the temperature, the stability of the floor, potential danger spots, attractions, and the like, according to the nature of his perceptual capacities. He looks over the ground so that he can keep one foot on it.

Human subjects do the same. They "look the situation over," noting

how high the ceiling is, where the windows and lights are, where the experimenter is, what the social climate may be, and where and when the next move may originate. Naturally certain objects stand out as figure, according to the subject's expectations, his perceptual and social fields. Objects available for manipulation are perceived and interpreted according to previous experience, interests, and the like. It is likely that the first survey is a sort of noncommittal exploration. The subject looks for possible traps, possible sources of excitement, trouble, or reward, without committing himself to any course of action prematurely. The extreme case—which may be an ideal that never actually occurs—would be pure survey of the situation without any response.

If the problem is a mechanical puzzle in the form of rings to be separated or pieces to be assembled into a pattern, this preparatory survey of the puzzle—looking it over, turning it around, observing possibilities and limitations of manipulation, inspecting the parts, etc. —is commonly noticed. If the problem involves relationships between words and other abstractions, the thinker may rehearse the meanings of the words, their connotations, and their relationships before producing any new verbal responses, but the time required for the preparation is usually short and the process is not externally observable.

That some things are eliminated as irrelevant or unprofitable is evident from comments during problem solving: "No use trying that again"; "This has nothing to do with it." At the same time attack may be centered on some part of the problem situation that is perceived as the center of the difficulty. This was called "locus analysis" by Ruger (*1910*) and "penetration" by Duncker (*1926*). Many other observers have used such metaphors as "getting to the heart of the problem."

If the problem involves verbal or arithmetical relations, a similar organization of the data probably takes place, a cognitive organization, it may be called. It seems likely that previous learning is more important when the problem is verbal or conceptual than when it is spatial or geometrical.

The result of this preparatory survey is that the problem is structured; the material at hand is organized. This structure of the problem situation becomes the effective environment within which productive thought operates. The thinker is now set to respond, not to the whole objective world and everything in it, but to the problem situation as he has organized it. The problem takes shape.

Naturally, problems that depend heavily on this organization of the problem material have been the favorites of the Gestalt psychologists, rather than problems that depend chiefly on the finding of new data by exploring the objective environment or one's memory. Six of the seven chapters of Wertheimer's (*1945*) essay, *Productive Thinking,* describe such problems. As Wertheimer puts it: "Central to the theory is the transition from piecemeal aggregation, superficial structure, to the objectively better or adequate structure. . . . Often, of course, when the data are adequately structured in this sense the solution falls into place."

Obviously the initial survey of the problem is of critical importance because it results in a set toward some aspects of the problem to the exclusion of others. The subsequent production process is channeled one way rather than another. If production following the initial preparation is unrewarding, the problem may be restructured or reformulated. Thus the preparation becomes a minor problem in itself, and alternative formulations of the problem are produced. For this reason it has been necessary to emphasize the wheels-within-wheels nature of the thought process. The structuring of the data is a subproblem, the solution to which formulates the next subproblem. Duncker (*1945*) puts it as follows: "The final form of a solution is typically attained by way of mediating phases of the process, of which each one, in retrospect, possesses the character of a solution, and, in prospect, that of a problem."

The conceptual scheme of this book tries to encompass the merits of the Gestalt emphasis on perceptual or cognitive organization and the functionalist emphasis on activity. The weakness of the concept of preparatory set, as ordinarily used, is the specification of the stimulus. Gestalt psychology has shown that the stimulus is perceived within a background or stimulus field and that the organization of this stimulus field should not be overlooked. The weakness of the Gestalt emphasis on structural factors is that the activity of the observer may be left out. *Keine Gestalt ohne Gestalter.* A broad view of the preparatory set repairs these two weaknesses. The thinker surveys the situation and prepares to respond to the stimulus field *as he organizes it.* The specification of the stimulus depends on both structural and functional factors.

SPECIFYING THE SOLUTION

As I was going to St. Ives, I met a man who had nine wives. And each wife had nine cats, and each cat had nine kittens. How many were going to St. Ives?

The other side of the preparation, complementary to the organization of the data, is the specification of the required solution. Failure to solve the above conundrum is due to failure to note what is required. Children often fail arithmetic problems because, as they become involved in the computations, they lose sight of their goal. For this reason teachers often insist that after their students write down what is given, they write down what is required. We can picture the thinker as prepared to find a certain solution response to a certain stimulus pattern. Productive thought is guided by this set, bridging the gap from either end or from both ends at once.

Perhaps in simple animals like rats the survey of the problem is not separate from the specification of the goal as it is in the more diffentiated human beings. Nevertheless, those interested in animal learning have been able to emphasize one side or the other in their theoretical formulations. Those influenced by Gestalt theory begin with the survey of the situation, preferring those problems in which the whole situation can be surveyed, and they try to state how the solution attempts reorganize the perceived situation. Those who favor response learning begin at the other end, with the goal response, and work backward, attempting to state how the goal response gets connected to its complex antecedents. Necessarily they prefer problems like the maze, in which the animal can somehow reach the goal before the problem is completely mastered.

In the complex problems of human concern the goal or solution may not be so clearly specified that the thinker can begin at once to produce tentative solutions. The preparation for thought may therefore include an analysis of the goal, making it more explicit.

It is difficult, of course, watching anyone approach a problem, to see what his preparation consists of. Reid (*1951*) attempted to demonstrate the importance of specification of the goal, not by observing his subjects perform this function, but by performing some of it for them, i.e., by experimental variation of the conditions. He had college students work construction problems with matchsticks and solid objects. Experimental groups received a standard series of aids designed

to "explicate the goal," while control groups received a parallel series of aids without explicatory content. For example, one aid for the experimental groups was as follows:

"It means that each match must form two triangle sides. If that is your original triangle there [pointing with finger], each of those three matches is already existing in one triangle. How can you put each of those three matches into two triangles at the same time?"

The corresponding aid for the control groups was as follows:

"People often find this problem rather difficult."

Such explication was helpful. The experimental groups solved significantly more problems than the control groups. There is no doubt that these statements by the experimenter were effective aids, and it seems likely that their effect was due to explication of the goal. At least it was a hypothesis about explication of the goal that framed the statements.

CONDITIONS THAT DETERMINE THE SET FOR PROBLEM SOLVING

If the above description of preparation for problem solving is approximately adequate, it is important to know what influences the outcome of this preparatory activity. What determines how the problem is formulated and how the goal is specified? We have noted in Chapter 3 that the set for simple activities may be prepared in advance by a motive, by social interaction, and by previous experience. The set for solving problems of a more complex sort is somewhat more elaborate but is established in the same ways.

MOTIVATION. In most animal experiments the solution or goal response is not specified, at least not at first. The motivation prepares the animal, e.g., for food, by sensitizing the perceptual mechanisms and lowering the threshold for discharge of a biologically prearranged response or consummatory mechanism. Prior to experience in the problem situation, however, the goal is not sharply specified. The consummatory response will be evoked indiscriminately by a wide range of food objects wherever encountered. After some experience with the reward many animals will discriminate between an expected and an unexpected goal object (Tinklepaugh, *1928;* Elliott, *1928*). It seems reasonable to assume that the animals do specify the goal-in-this-situation and develop an expectancy for it.

Another effect of the motivation is to raise the general activity level, increasing the readiness to act.

SOCIAL INTERACTION. The above considerations apply to human

problem solving, but, since the motivation is usually social, the preparation is complicated by social interaction. In the laboratory and classroom the problem material is given by the experimenter, teacher, textbook, or examination question, and the goal is specified with varying degrees of explicitness. The thinker also makes certain assumptions from the atmosphere of the situation. He assumes usually that the experimenter is honest; that one is not supposed to break anything, etc. A problem used by Bulbrook (*1932*) could only be solved if some beads were broken, and this assumption was a major source of difficulty. Under other social conditions, e.g., alone on a desert island, this assumption probably would not be made.

The subject's set can be varied, of course, by explicit instructions from another person, and different instructions lead to different degrees of problem-solving success. Several experimenters have changed a thinker's approach to a problem by various hints, or by moving objects in front of the thinker. Teachers often structure a problem for their students, helping them formulate the problem and specify the goal.

Marks's (*1951*) problem-solving experiment demonstrates how the experimenter's activities can influence the way the subjects attack the problem. The problem was to find the source of errors made when square roots were computed on a calculating machine with the help of a table of divisors. (The errors were due to false entries in the table of divisors.) Matched groups of college students were compared in respect to the effects of three activities of the experimenter.

One group received lectures, three weeks before the experiment, on how to solve problems. As part of a regular psychology class the students were instructed in the analysis of a problem into its elements, the determination of the requirements for solution, the formulation and testing of hypotheses, etc. Examples were given. Of this group of 22, 10 solved the problem. Of a control group of 20, only 5 were successful. The difference is not statistically significant, but it offers a ray of hope to discouraged psychology teachers.

For another group the experimenter analyzed the problem. He furnished them a typed list of four elements—the operator, the calculating machine, the method, and the table of divisors—telling them that the error could be attributed to one of the four. This list was not very helpful, however, because a control group got practically the same percentage of solutions.

The experimenter aided another group by saying at intervals: "I want you to ask yourself, 'What are the elements of my problem? What are the only ways in which a mistake could occur?'" If the subject strayed from the analytical attitude engendered by the questions, he was brought back to it by repetition. If he mentioned some elements of the problem, he was encouraged to test them and search for more. But the experimenter did not name elements. This type of aid was distinctly effective. All 11 of this group solved the problem, as compared with a control group in which only 5 out of 30 were successful.

From this experiment we may conclude that someone else's analysis of a problem is not much good. The thinker may listen to it, but he does not necessarily accept and use it. Teaching a general analytical approach may, however, transfer to a new problem. But the best thing the experimenter did was to force the subjects to do their own analyzing.

Another interesting comparison in Marks's experiment was between personal and impersonal situations. One group of subjects worked some problems and then was told to find the cause of the errors. The other group was given the same problems in the form of a typed narrative about one John Jones. The second or "impersonal" group achieved significantly more solutions. Presumably because they were less personally involved, they were better able to maintain an analytical attitude.

Previous Experience. Experiments on transfer of training have demonstrated that material learned in one situation may later be used in a similar situation—for better or worse. By the same principle of transfer the thinker may approach a problem with a preparation carried over from previous experience with similar problems—for better or worse.

In some complicated unfamiliar problems the thinker establishes his set as he goes along—the task-induced set—and changes it from time to time. In these cases the preparation process is not clearly marked off from the production process. Hence the experiments on this topic are more conveniently reviewed in the next chapter.

CREATIVE THOUGHT

What is usually called creative thought differs only in degree from the forms of thought previously described, but this difference is largely

a difference in preparation. Hence the essential characteristics of creative thought must be identified and preparation for it must be described.

THE MARK OF CREATIVE THOUGHT. Suppose we ask a friend to write a poem celebrating the opening of the fishing season. Let us ask him further to work out the most convenient method of transporting a party to the highly recommended fishing grounds north of Gull Island in the middle of Long Lake. Now why is it that performance of the first assignment will usually be called a creative act and performance of the second probably will not be thus labeled?

Most attempts to differentiate creative thought from more prosaic intellectual activities come sooner or later to the criterion of novelty. The novelty or originality of the product has frequently been offered as the defining mark of creative thought. But closer examination will show that we do not consider the product so much as the process. If two architects draw up the same plans, and we know that they worked independently, we do not consider the work of either any the less creative for that reason alone. If a geometry teacher gives his students an "original" to do and one student discovers the same solution that Euclid discovered, he will not therefore be considered lacking in originality. It is easy to find illustrations of an opposite sort as well. In some fields of research, e.g., industrial chemistry, one may develop a new material, like a plastic, without great creative thought merely by following techniques familiar to all well-educated chemists. The same may be said for new developments in plant breeding and the construction of psychological tests. The real criterion for creative thought seems to be, not whether society considers the product new, but whether the process of producing it is original.

If we turn then to an attempt to describe the process of creative thought, the criterion of novelty is still useful, but only in an instrumental way. The inventor and the scientist, for example, are not directly concerned with novelty, but they are working on problems for which routine procedures are not adequate, hence the criterion of novelty comes in a roundabout way into the specifications of the solution. The poet, novelist, sculptor, and architect are more explicitly concerned with novelty; they are prepared to reject the clichés. Hence the outsider, the critic, or society at large uses the criterion of originality of the product to infer originality of the process.

Otherwise creative thought is carried on much as are other problem-

solving enterprises. The problem is surveyed, the goal is previewed, and from these two sources an incomplete pattern is prepared which guides the thinker's productive activities. As in other varieties of thought, the preparation may change as activities continue, and new data are brought into view, or—especially in the case of the artistic thinkers—as new moods are aroused.

If this preliminary examination is correct, the defining characteristics of creative thinking are in the preparation process. A creative problem is one for which the thinker cannot well prepare. Asked to write a poem about the opening of the fishing season, he surveys the material at hand; but all the words he knows, all the ideas, metaphors, and high hopes he has ever dreamed of, as numerous and as diverse as the fauna of Gull Lake, are pertinent to his task. He identifies his goal, but any kind of poem can be as good as any other kind. His preparation can only be very general, not nearly as specific as preparation for planning a method of transportation.

Functionally, then, the essence of creative thought is a negative one. No known solution is sought and previous experience is not in itself sufficient. The goal is not fully specified and the thinker does not know which aspects of the problem situation are to be used. The preparation is loose, flexible; no helpful preparatory set can be organized. Usually preparatory sets are established and then abandoned as the responses evoked with their aid are rejected. In general the creative enterprise is freer, less directed, than is problem solving directed toward finding the one right answer, even though the difference is only one of degree.

Evidence of Preparation. If we must emphasize the freedom and flexibility of creative thought, why speak of preparation at all? Perhaps genuine creative thought goes at once into a productive phase without any preparation. The evidence is only subjective, but many creative thinkers have written quite clearly and consistently of a preparation period preceding the production of creative ideas.[1] In fact, one of the most frequent recommendations that distinguished scientists and inventors make to beginners is to load up with the facts, to study the problem from several angles, so as to see clearly what is given and what is required. Just because the preparation cannot be specific, the best preparation is broad and deep. Poets and artists often speak of immersing themselves in a mood before beginning active work. Novelists

[1] Translations from the testimony of Helmholtz and Poincaré are available in Woodworth *(1938)*. Other illustrations may be found in Hutchinson *(1949)* and Ghiselin *(1952)*.

often study their locales and their characters for some time before sitting down to their typewriters. It was subjective comments of this sort that led Wallas (*1926*) to identify and name the preparation period.

Attempting to secure more tangible evidence, Patrick (*1935*) in her study of poets assumes that the preparation is changing until the first line of poetry is formulated. Therefore she takes, as her sign of preparation, simply the thought changes, modifications of thought sufficient to cause it to form a new sentence. Thought changes, as thus described, are much more common in the first quarter of a creative period than in any other. The difficulties with this technique have been discussed in Chapter 2, but an excerpt from one of Patrick's protocols (Poet C) will illustrate the character of the evidence:

> Here I am presented with a bum picture and told to write a poem, but certain things come to mind with the mountain peaks at the top and the gloomy cavern below and the great chasm that goes clear down to the center of the earth might form a poem. A purely imaginary landscape is my first impulse. It should be understood these are peculiar circumstances. I tried something about the gloominess and the lower part of it and reaching up to the top but discarded it. I might with time work the thing out, but not now. I confound the spiritual with the mountain tops; I do not say the mountain tops are confounded with the spiritual. Literature is full of that. Now I think what it is. The best thing is that cavern. That is the thing that goes down. I take it you want a serious poem. Well, suppose, suppose I begin.
>
> a. There was a cavern in Xanadu.
> b. Like this before me on the painted page. No idea how to go on with derivatives, but never mind (pause). Well I might go on. It won't be any good, but never mind.
> c. and d. Where water wove and weltered from age to mystic age. Kind of bum, but maybe do something afterwards. Had my idea. The feet are all wrong.
>
> There was a cavern in Xanadu . . .
>
> And so on.

By the very nature of creative thought the initial preparation is not usually retained. Certainly the initial preparation is quite vague and general. It is here, as much as anywhere, that thought shows its wheels-within-wheels character. It may begin merely with a mood or very general specification of the desired solution, but within this general orientation a more specific preparatory set may be established. Several

of Patrick's poets, for example, left blanks in their first drafts; then returned and, with a more definite set, were able to fill the blanks. One poet wrote:

> The very small heart of a very small fool
> is _____ a very small pain.

This poet later returned and wrote "aware of" in the blank.

> PERSONAL PROBLEMS
> How can I get a B in Econ. 310?
> How can I get a date with Leslie?
> How can I stop biting my fingernails?
> What should I do to get a job? or a raise?
> Do I really believe in God?

All problems are personal problems in a way since people are always personally interested in success rather than failure. But people take some problems more personally than others. Hence the difference between personal and impersonal problems is one of degree—degree of personal or ego involvement. If the thinker considers the problem remote or academic, the outcome being of no special consequence to him, we say that he is not personally interested or not highly ego involved. Many of the laboratory problems thus far discussed are of this sort. The problem becomes more personal when failure is a threat to the ego, or when the outcome has a bearing on the thinker's self-esteem or social status. A textbook problem in addition and subtraction may be quite impersonal to a fifth-grade boy, but a problem of saving money for a bicycle involving the same numbers and the same operations would be taken personally.

Degree of ego involvement in a task is not necessarily constant. A problem that is initially impersonal may acquire some personal significance if the thinker experiences success and failure while working on it or if he develops a preference for one kind of solution rather than another. For experimental purposes psychologists have found it possible to manipulate ego involvement by varying the instructions. A series of problems may be presented to a group of subjects merely because the experimenter wants to find out how much time they require. Or the same series may be presented as an intelligence test, to determine the IQ's of the subjects.

The motivation for personal problems is highly varied, for obstacles may arise to block any motive. People tell psychologists most often about the problems they do not solve, but if a thorough census of all kinds of personal problems were taken, it would probably show that most people do solve most of their personal problems. One reason for this is that the personal involvement in the problem usually increases motivation. Problems of the psychological laboratory and classroom may be a challenge, to be attacked with a burst of intellectual activity, then forgotten. But the preparatory set for personal problems lasts longer. A person works on his problem for a while, then he goes to a movie. He struggles with it again on the way home, then goes to bed and perhaps dreams of it. He may take it up again in the morning or at least in the next spare moment. He may even go to an expert for help. Such persistence facilitates the development of varied approaches to the problem and therefore increases the chances of success.

Personal problems are also less abstract than impersonal problems. Different approaches to the problem have more meaning, and inferences about consequences of these approaches strike home. It is easier to get the lay of the land when one is in the middle of it. The preparation for personal problems includes, on the static side as a result of past experience, memories, likes and dislikes, techniques of adjustment, sets, and the like. Some of these transfer automatically to the new problem and are included in the survey of the problem. The relevance of other material from the past may be noted only as problem solving proceeds.

For these reasons textbooks of educational psychology commonly recommend that problems be made personal. Johnny can be taught not merely to add two and two but he can be asked how many apples he would have if he had two and was given two more. College students can study problems in a psychology textbook, but the same problems can perhaps be presented to them as their own personal problems.

On the other hand personal problems are not always easier than impersonal ones, as in Marks's experiment mentioned on page 164. There are several possible reasons for this inhibiting effect of personal involvement.

In the first place personal problems are likely to arouse more emotional excitement. Emotion has its principal effect on the production process and will therefore be considered in the next chapter.

The personal involvement may have a critical effect on the prepara-

tion process, both on the survey of the problem situation and on the specification of the goal. In formulating the problem, not all the material at hand is always used. Some of the material pertinent to the problem may be suppressed or repressed, and perceived relationships may be distorted. In defining the goal similar phenomena may occur. The thinker may start with the aim of solving some problem, but may later veer off toward defense of his first solution attempt, or of his formulation of the problem. As he nears the goal he may anticipate distasteful consequences of success and therefore substitute some other goal. Anxieties about outcomes are distracting, as compared with an impersonal interest in the problem for its own sake. Goals may be incompatible. One cannot spend the same dollar for two commodities or sin and feel virtuous at the same time. Personal problems often involve redefining of goals and fitting temporary goals into the long-term pattern of goals before the production of solution attempts can get under way. Hence problem solving may begin with or acquire an incomplete or inadequate preparation. With this preparation the production process may proceed as for impersonal problems, because the preparation has in effect rendered the problem impersonal.

The extreme case of personal involvement hindering problem solving is the case of the neurotic who is willing to spend money and time on his personal problems, but formulates his situation and goal in such a way that solution is impossible. Many neurotics seem to be capable of "thinking through" their difficulties if the therapist helps them, in the preparatory phase and during the course of therapy, by arranging conditions that are better for "thinking through" than those the neurotic has established for himself.

INDIVIDUAL DIFFERENCES IN PREPARATION

The preparation for the discrimination experiment is standardized so that individual differences will be minimal. Uncoöperative subjects, i.e., those who do not adopt the required set, are eliminated. Thus the individual differences that occur in the discrimination experiment are of more concern to the psychology of the sensory and perceptual process than to the psychology of the thought processes.

FREE ASSOCIATION

Individual differences in free association have been studied for many years, principally for practical purposes. In the terminology of today

the free association test would be called an unstructured, or perhaps a projective, test since the subject is relatively unrestricted in his responses. We have seen that the responses come from a pool or supply of words and that they are controlled by the subject's set for the experiment, which is not unrelated to his set for other social situations. One would expect, therefore, that people would differ widely from one another in the responses that they give, both because of the words that are available to them (the static aspect of the preparation) and because of the sets with which they approach the test (the dynamic aspect of the preparation).

The free association test was invented long before our present concept of test reliability was developed, but Karwoski and Berthold (*1945*) have recently approached that fundamental problem by giving a list of 59 words to 50 students and repeating it 40 days later. The responses were classified into seven classes (see page 148) by the same investigators each time, and frequencies of words in each class on the two occasions were correlated. For the seven classes the correlations ranged from .61 to .92. When four classes based on response similarity are combined, the reliability of this score becomes .92; and when two classes based on contrast are combined, the reliability of this score is also .92. From the intercorrelations it appears that the two factors of similarity and contrast account for most of the variance. These reliability coefficients are high enough to justify further use of the scores obtained in this way. An illustration of the possibilities of the use of such scores is an early study by Murphy (*1917*). He found that among university students and faculty those in the science department tended more toward coördinates and contrasts, while the literary group favored completion and predication. More recently Foley and MacMillan (*1943*) gave free association tests to students of law and medicine and found significant differences in their responses. To the word "administer," for example, the law students gave such response words as administrator, business, decreed, direct, estate, execute, and govern; while the medical students gave such words as adenoids, aid, anesthetic, attend, doctor, dosage, drug, and first aid.

The older experiments on word association used classification procedures of unknown reliability. But if the same procedure is used on two samples of words, errors are probably constant and the differences are likely to be genuine. From this older literature it appears that some subjects, especially children and adults of limited education, take

the experiment as a test of ability and are thus prepared to display their vocabulary knowledge, to tell what the word means. They give the defining or general identification response more often than average adults. Educated adults give more coördinate and contrast responses, presumably because they do not need to impress the experimenter, but, willing to coöperate, they give an easy, reasonable response.

Practical use of the free association test is based in part on the assumption that the subject's preparation for, or attitude toward, life in general overlaps his preparation for the stimulus word, and in part on the assumption that the emotion aroused by critical words will disrupt the preparation and yield a response (verbal, gestural, physiological), of a personal or evaluative nature, which can be related in some way to the subject's personality.

RESTRICTED ASSOCIATION

The uncertainties of classification of response words have led to the construction of the multiple-choice form of the association test. The instructions are the same, but after the stimulus word the subject is shown two or more likely response words. This procedure has been used by Goodenough (*1942*) and Terman and Miles (*1936*), among others, with reasonably good reliability. One of the advantages of the multiple-choice form is that the response words may be selected, on the basis of research, to correlate with a criterion, or with contrasting groups. Terman and Miles used this procedure in their test of masculinity-femininity, making use of pictures as well as words. Since the stimulus words are the same for all subjects, the difference between men and women must be a difference in what they bring to the task, i.e., the preparation.

If the analysis of the earlier part of this chapter is correct, association tests may be so highly restricted that they become tests of ability rather than personality or interest. If one of the choices is right and the others wrong, and the set is standardized by the instructions, the test becomes a test of the static aspect of the preparation, based on past learning.

PROBLEM SOLVING

There is plenty of evidence that remote preparation in the form of general knowledge or supply of information correlates with problem-solving achievement in any problem area. But when we ask about preparation for any specific problem, it appears that tests for ability to

prepare for a problem, as differentiated from fluency of production of solutions or ability to criticize solutions, have not been devised. Factor analyses, for example, have not been directed toward this sort of analysis. Educational psychologists interested in testing for meaning or understanding have come closer to the preparation phase. For instance, in an N.S.S.E. *Yearbook* (*1946*), Part II, "The Measurement of Understanding," Findley and Scates have a section on measuring the "ability to recognize the relevancy and sufficiency of data," and give the following, taken from Mathematics Progress Tests prepared for use in the schools of New York State, as an example.

Read each problem below and the facts which follow it. Decide what facts you need to solve the problem. After each fact you need, write YES. After each fact you do not need, write NO. Then answer the question "Do you need more information?" by writing YES if you need more information, or NO if you do not.

How far did Mr. Day travel on his trip?
1. He left Monday morning and returned Wednesday evening.
2. He drove 87 miles on Monday.
3. He drove 42 miles on Tuesday.
4. Do you need more information?

In the same yearbook Anderson, Forsyth, and Morse in the chapter on the social studies devote some attention to test items which test defining the problem, using appropriate sources effectively, selecting and organizing pertinent information about a given problem. The problems under consideration could be called creative problems, as in writing an essay, and judgment problems, as in evaluating testimony.

Directions: Number sources listed in the order of value in studying about the living conditions of ancient Romans.
An epic poem written by a Roman which mentions some family customs.
Excavations of the buried Roman city of Pompeii.
A novel about the fall of Pompeii.

Naturally one would like to know how success on these items correlates with success on other kinds of items, with reading ability, and with general intelligence, and measures of achievement in these fields. Such analytical data would help also in the problem of analyzing the thought processes.

Correlational investigation might reveal a special synthesizing or organizing ability, necessary for assembling pertinent data at hand,

into a pattern or search model useful for problem solving. Presumably such an ability would not be required for restricted association, wherein the search model is given by the instructions, but would be required for problem solving, creative thought, and judgment. If there is such an ability, it has been noted in popular language by such phrases as "he cuts through to the heart of the problem," "he simply does not grasp the problem," and "he is a good troubleshooter."

Perhaps there are important individual differences in preparation which would come under the heading of personality rather than ability. If we could separate the preparation phase from the others and measure the amount of time different people spend on each, we might find a trait of some consequence in daily life. At least we know that some people are characterized as prone to give hasty solutions or snap judgments, while others like to "be prepared" before they set off on an intellectual adventure.

Some evidence on this point is available in a recent study of instruction times by Wells (*1950*). College students, taking ability tests individually, were allowed to prepare for each by reading the printed instructions and signifying when ready to begin. Time of preparation for each test was recorded. These instruction or preparation times were correlated with each other and with the test scores.

In four groups of subjects taking different tests, the results were the same. Results for the largest group, of 200 subjects, would be the most significant. In this group instruction times for the six subtests of the Army Alpha were intercorrelated, yielding 15 correlation coefficients. These coefficients ranged from .33 to .66 with a median of .49. Evidently those who get ready for one test quickly also get ready for another quickly.

This common factor in the preparation times is probably made up of ability and personality variables, corresponding to the static and dynamic aspects of the preparation. The weight of the ability variable can be estimated by the correlation between preparation times and test scores. Such correlations are negative because long preparation times go with low ability. In this sample of 200 students the six correlations range from —.13 to —.34, with a median of —.26. Probably reading ability is involved in these correlations because the correlations between preparation time and test score are slightly higher for the verbal subtest than for the numerical subtests. The difference between the median correlations of .49 and .26, however, indicates that the common

factor in preparation times is not only a general ability but in part a personality factor that might be called "willingness to take time to be prepared." Wells mentions the terms impulsiveness and compulsiveness. A short self-descriptive term would be *forethoughtfulness*. Wells's technique of timing the instruction period suggests a method that could be adapted to preparation for all kinds of laboratory and test assignments.

A little evidence on preparation time is available from studies of 57 men between 20 and 70 years of age carried out by Welford (*1951*) and his students. The task was a kind of serial judgment or inspection task. Aluminum blocks came into view, one at a time, down a runway. Each was to be judged up to standard, in respect of width and design, or rejected. When the speed of presentation was under the subjects' control, older subjects proceeded more slowly than the young, especially at first, and were somewhat more accurate at first. On this task and on more complicated problems, "older subjects often required the experimenter to go over the instructions more than once, and frequently rehearsed parts of them aloud before pronouncing themselves ready to proceed with the task."

REFERENCES

Bulbrook, M. E., 1932. An experimental inquiry into the existence and nature of insight, *Am. J. Psychol., 44*:409–453.

Duncker, K., 1926. A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems), *J. Genet. Psychol., 33*:642–708.

Duncker, K., 1945. *On problem-solving,* Psychol. Monogr., 58, No. 270.

Elliott, M. H., 1928. The effect of change of reward on the maze performance of rats, *Univ. California Publ. Psychol., 4*:19–30.

Foley, J. P., Jr., and MacMillan, Z. L., 1943. Mediated generalization and the interpretation of verbal behavior: V. "Free association" as related to differences in professional training, *J. Exper. Psychol., 33*:299–310.

Ghiselin, B., 1952. *The creative process, a symposium,* Berkeley, University of California Press.

Goodenough, F. L., 1942. The use of free association in the objective measurement of personality, in *Studies in personality,* ed. Q. McNemar and M. A. Merrill, New York, McGraw-Hill Book Co.

Herrick, C. J., 1939. *Awareness in the inductive solution of problems using words as stimuli,* Arch. Psychol., No. 234.

Hutchinson, E. D., 1949. *How to think creatively,* New York, Abingdon-Cokesbury Press.

Jersild, A. T., 1927. *Mental set and shift,* Arch. Psychol., No. 89.
Karwoski, T. F., and Berthold, F., Jr., 1945. Psychological studies in semantics: II. Reliability of free association tests, *J. Social Psychol., 22:*87–102.
Karwoski, T. F., Gramlich, F. W., and Arnott, P., 1944. Psychological studies in semantics: I. Free association reactions to words, drawings, and objects, *J. Social Psychol., 20:*233–247.
Karwoski, T. F., and Schachter, J., 1948. Psychological studies in semantics: III. Reaction times for similarity and difference, *J. Social Psychol., 28:*103–120.
Marks, M. R., 1951. Problem solving as a function of the situation, *J. Exper. Psychol., 41:*74–80.
Murphy, G., 1917. An experimental study of literary versus scientific types, *Am. J. Psychol., 28:*238–263.
National Society for the Study of Education, 1946. *45th yearbook,* Pt I, The measurement of understanding, Washington, D.C., the Society.
Patrick, C., 1935. *Creative thought in poets,* Arch. Psychol., No. 178.
Razran, G. H., 1949. Some psychological factors in the generalization of salivary conditioning to verbal stimuli, *Am. J. Psychol., 62:*247–256.
Reid, J. W., 1951. An experimental study of "analysis of the goal" in problem-solving, *J. Gen. Psychol., 44:*51–69.
Ruger, H., 1910. *The psychology of efficiency,* Arch. Psychol., No. 15.
Terman, L. M., and Miles, C. C., 1936. *Sex and personality: studies in masculinity and femininity,* New York, McGraw-Hill Book Co.
Tinklepaugh, O. L., 1928. An experimental study of representative factors in monkeys, *J. Comp. Psychol., 8:*197–236.
Wallas, G., 1926. *The art of thought,* New York, Harcourt, Brace & Co.
Welford, A. T., 1951. *Skill and age,* New York, Oxford University Press.
Wells, F. L., 1911. Practice effects in free association, *Am. J. Psychol., 22:*1–13.
Wells, F. L., 1950. Instruction time in certain multiple-choice tests; Cases XCVI–CII, *J. Genet. Psychol., 77:*267–281.
Wertheimer, M., 1945. *Productive thinking,* New York, Harper & Brothers.
Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt & Co.

Chapter 7

The Production Process: Simple Responses

If at first you don't succeed, try, try again.

NEARLY all classifications of the thought processes, whatever the method of classification, have included, in some way or other, the production of varied responses. Dewey (*1910*), for example, speaks of "suggestion of possible solution." Wallas (*1926*) emphasized the particularly enlightening suggestion, happy thought, or "illumination," coming after many unilluminating thoughts. Following Wallas, Patrick recorded as evidence of illumination the first drafting of lines of poetry (*1935*) and the first sketching of general shapes (*1937*). In experiments on problem solving the observers have recorded number and kind of movements made of solutions attempted and, in the case of human subjects, number and kind of verbal responses.

The adjective "productive" has sometimes been used in an evaluative sense, as in "productive thought," to refer to good thinking or thinking that reaches good conclusions—as contrasted with unproductive thinking. In the present system of concepts the production process, or the productive phase of thought, merely designates the emission of responses in a problematic situation, whether these responses be simple or complex, good or bad, verbal, manual, or locomotive. We have already seen how preparation can be differentiated from production logically and in some cases experimentally. We shall attempt in later chapters to differentiate judgment from both preparation and production.

The production process might be called the center of the psychology of the higher mental processes. In order to see the center more clearly we have cut off the preparation process and the judgment process for separate treatment. Nevertheless the information on this topic is voluminous and heterogeneous. Production of relatively simple responses, especially words, will be considered in this chapter. The next will take up the production of more complex patterns.

PRODUCTION OF SINGLE WORDS

The free-association experiment has given us much information about preparation, in both static and dynamic aspects, most of which has been summarized in the two preceding chapters. It tells us little about production of words, however, because, once preparation is established, the response word is produced quickly, then production is finished. All we learn about production from the free-association experiment concerns time of production of the response word after the stimulus word is given.

Experimental psychologists have long known that those responses which are made most frequently in free association are made most rapidly. A number of subjects are presented with stimulus words, one at a time, from a standard list, and the frequencies of the response words tabulated. Many response words occur only once, of course, but some stimulus words evoke the same response word from a large percentage of the subjects, e.g., slow-fast. When response time is plotted on percent frequency (or communality), an inverse curvilinear relationship appears which is called Marbe's law. Results from several experiments have been graphed by Woodworth (*1938*).

Schlosberg and Heineman (*1950*) have recently repeated this experiment with 204 college students and find that using the logarithm of time straightens the curve. Hence the relationship or law is approximately of the form $\log Y = c - aX$, where Y is response time, X is percent frequency, and c and a are constants.

It has often been an occasion for surprise that restricted association is not slower than free association—though the early experiments left many variables uncontrolled and the difference was not thoroughly established. Recently Baker and Elliott (*1948*) have made the proper comparisons, from which it appears that restricted association is definitely faster than free association, other things being equal. The facts

sound less surprising when rephrased: a well-prepared response is produced faster than one less well prepared.

A task similar to restricted association is the memory experiment that asks for production (or reproduction or recall) of words previously learned. Emphasis is usually centered on the conditions of acquisition of the words and on any activity that may be interpolated between the acquisition and the recall (see chart on page 117). Conditions of recall *per se* are more important for the present chapter because, if the first attempt at recall is unrewarding, the subject may be said to have a problem. He may report that he is trying to "think of" the word or that he is "searching his memory."

When the subject does try to think of a word, the errors or attempted solutions give a clue to the production process. Often these errors are related in some way to the wanted word. If the wanted word is port, for example, harbor, sherry, or part may be produced. (See Woodworth, *1938,* p. 37.) There may be some spread or generalization along gradients of similarity, e.g., of sound, sense, or spelling, or the relationship may be one of simple fortuitous association.

Instead of waiting for errors to occur, Underwood and Hughes (*1950*) studied errors in recall experimentally, by a modification of the paired-associates experiment. They had college students learn pairs made up of a nonsense syllable and an adjective, e.g., xan-musty, and tested them a week later by exposing only the nonsense syllable. The subjects were urged to respond even if they could not remember at all.

Most of the responses were correct and thus useless for present purposes. The erroneous responses were classified by a group of judges as well as possible in respect of their relation to the correct response words. Of the meaningful relations, synonyms were most frequent, homonyms next, and antonyms least frequent. In most cases the judges were unable to determine the basis for relationship between the word produced and the word memorized a week earlier. But the number of apparent relationships was much greater than for a control group that did not do the previous memorizing.

This experiment demonstrates that some such process as generalization, or mediated association (see page 129), influences the production of words in memory experiments as in association experiments. The evidence for a gradient or dimension of similarity is less clear-cut, because most of the response words were judged either highly related or unrelated to the correct word.

An interesting question now arises: When does the generalization take place? The usual assumption is that generalization or association occurs at the time of the original acquisition of the response, strengthening related responses. If, however, we extend Razran's "subsequent test" (see page 129) interpretation of semantic conditioning to the memory experiment, it would appear that generalization may occur also at the time of recall. In other words, recall may be an active or production process. Razran's ingenious trick of using controlled association to establish a specific set, e.g., to name supraordinates, could be used just before the recall and just before the original acquisition to find out which placement produces more supraordinate errors of recall.

In these experiments on single words the production process runs off rapidly. The subject produces a word, right or wrong, and he is finished. Many subjects do not, perhaps cannot, report the words they think of before they deliver the response word. If mediated generalization or something similar occurs, some of the productive activity may be obscured. It is necessary to bring the productive process into observation as much as possible, and the next type of experiment promises to do so.

CONTINUOUS PRODUCTION OF WORDS

The geneticist finds it profitable to study fast-breeding organisms, like the fruit fly, which give him abundant material for analysis. There is a corresponding advantage for the psychology of thought in arranging tasks in which responses are produced copiously, preparation remaining constant. Such a task is the one historically called "continuous association." An example of this task is the item on the Stanford-Binet Test at the 10-year level which asks the child to name words, any words at all, for three minutes. In spite of the advantages of preserving historical names, it is stretching the name "association" too far to call this continuous association. Emission or production of words is a more appropriate name.

As in the case of single-word association, continuous production of words may be restricted in a variety of ways. The subject may be asked to name automobiles, presidents, fruits, words ending in "tion," unpleasant people, and what not. If the restriction is not too severe, the flow of words will be copious, yielding plenty of responses for quantititive and qualitative study. The number of words produced in a standard time is taken as a measure of fluency. There is some ex-

perimental evidence on conditions affecting the output of words, and there have been many correlational studies of individual differences in fluency.

CONDITIONS AFFECTING FLUENCY

One would expect that some physical factors, such as drugs and loss of sleep, would affect the rate of production of words and other responses. In one study of 23 medical students, dexedrine sulfate increased fluency significantly (Welch, Diethelm, and Long, *1946*). In another study of eight young men maintained in acute vitamin-B deficiency for 23 days, there was no reduction in performance on fluency tests (Guetzkow and Brozek, *1946*).

Emotional conditions would also be expected to influence word fluency. Lantz (*1945*) investigated the effects of frustration on the performance of 9-year-old children on the Stanford-Binet Test and found that the item that was most disturbed by frustration was the one that asks them to "Name all the words you can think of."

This suggests that a cheerful mood would encourage fluency while emotional depression would block fluency. Among patients in mental hospitals manics are typically reported as talkative and depressed patients are typically nonproductive. Studman (*1935*) and Welch, Diethelm, and Long (*1946*) have taken the trouble to check this clinical observation by administering fluency tests to psychiatric patients. In the later study the elated patients of all classifications were much more productive than the nonelated. The highest scores were made by the elated manic-depressive patients and the lowest by the depressed patients. Data from correlational studies, to be described at the end of this chapter, tie in with this principle.

The old saying "bending the elbow loosens the tongue" is based on the common belief that alcohol increases fluency. Casual observations seem to confirm this principle, but under the usual conditions of observation the observer's accuracy is not beyond question. A carefully controlled study by Hartocollis (*1954*) attacked this question directly with the aid of four kinds of fluency tasks: (1) Fluency unrestricted: "Name all the words you can think of." (2) Fluency with slight restrictions: e.g., "Name words that begin with M." (3) Fluency with moderate restrictions: e.g., "Name trees." (4) Fluency restricted to one correct word: e.g., "Name a place or building for athletic exercise that begins with G." The subjects were 60 young men, with heavy drinkers

and nondrinkers eliminated. The experimental subjects got a drink that contained 38 cubic centimeters of ethyl alcohol per 100 pounds of body weight.

The alchohol increased output slightly on the third task, fluency with moderate restriction, but definitely reduced output on the other three tasks. Analysis of the words produced indicated that the increase on the third task was due to a lowering of standards. The control subjects held themselves to production of words that clearly conformed to the instructions while the experimental subjects were less critical. For example, when asked to name trees, the experimental subjects gave such questionable responses as "Christmas tree," "family tree," and "big tree." This relaxation of standards did not raise output on the first and second tasks because practically all words that might be produced were acceptable. Nor was relaxation of standards helpful on the fourth task because here only one word was correct. But when the task required a moderate degree of restriction, as in casual social conversation, relaxation of standards apparently was responsible for the small increase in productivity.

MATHEMATICAL ANALYSIS OF THE PRODUCTION OF WORDS

The score usually used in the study of fluency has been simply the number of words produced in a standard period of time. We may call it a raw score. A more analytical technique for investigation of the continuous production of words has been published by Bousfield and Sedgewick (*1944*). They asked college students: "Name all the cities in the U.S. you can," and "Name all the quadruped mammals you can." The subjects wrote these words down and drew lines at the end of every two minutes. By counting the words written in each two-minute interval and cumulating these, Bousfield and Sedgewick were able to draw the curves shown in Figure 7.1.

Mathematical analysis of these sequences of words promises to increase our understanding of the production process considerably. Bousfield and Sedgewick assume that the subject has a supply of words that fit the instructions and that the rate of production of words is proportional to the number of words remaining in reserve. Suppose, for example, that the subject knows the names of 100 cities. In the first two-minute period he may think of a fifth of these, or 20. That leaves 80 in reserve. If his rate of production remains constant, he will recall about a fifth of these, or 16, in the next two minutes. So his total is 36,

and 64 remain available in reserve. A fifth of 64 is 13, which would be the number of names emitted in the next two minutes. And so on. If we use n for the total number of words emitted, the number emitted in any two-minute period would be ⅕ of (100—n). To apply this to all cases we can let c stand for the number of words in the supply and m for the rate of depletion of this supply. The number emitted, then, in any particular time interval would be $m(c—n)$.

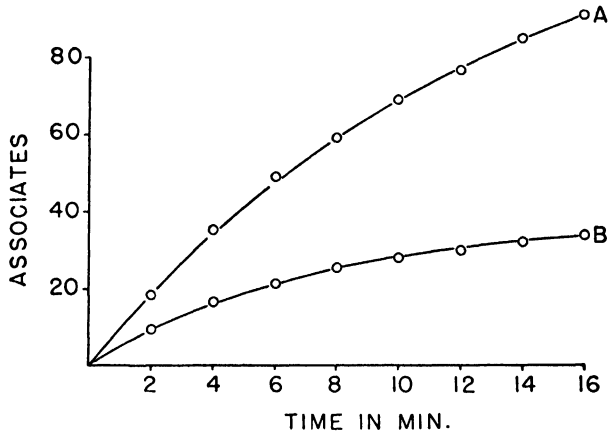


FIGURE 7.1. Continuous production of words. Plot A represents the average sequence obtained with instructions to write the names of fellow college students, and Plot B the average sequence with instructions to write animal names of two or more syllables. The circles indicate the means of the group data. The broken lines depict the approximate theoretical curves. For Plot A, $c = 128$ and $m = .0775$; for Plot B, $c = 37.3$ and $m = .1427$. From Bousfield and Sedgewick (*1944*).

Bousfield and Sedgewick plot their curves, not as rate of emission, but cumulatively, as total number of words produced as a function of elapsed time. In the above example n increased from 20 to 36 to 49, and so on. In the form in which Bousfield and Sedgewick plot their curves, $n = c\ (1 - e^{-mt})$. This equation, in which e is the base of the natural logarithms and t is elapsed time, is in the exponential form so commonly used in physics, chemistry, and physiology to describe the rate of approach to a limiting condition.

This theoretical equation certainly seems reasonable. The next question is: Does it agree with the facts? The answer is given in part by Figure 7.1, which shows that the theoretical curve does describe the data quite well in some cases at least. Thompson and Kepler (*1945*) and Johnson, Johnson, and Mark (*1951*) also were able to fit this curve to their data obtained in the same way. In general the theoretical equation described the production of words quite well, for both individual and group data, in spite of minor discrepancies.

As contrasted with the conventional method of studying continuous production, which yields only a raw score, the value of the Bousfield and Sedgewick technique is that it describes production of words in terms of two analytical scores. These two scores are obtained by fitting the production curves with a mathematical equation and calculating the two constants, c and m, representing respectively the supply of words and rate of depletion of the supply. That these two constants are not mere mathematical artifacts but do have rational psychological significance is shown by arranging conditions under which the constants take on different values. One is a study of overlearning and the other compares the production of pleasant and unpleasant words.

Overlearning, i.e., studying a list of names of automobiles for five minutes, increased the rate of naming automobiles (m for an unpracticed group of 9 students was 0.219; while for a group of 17 with practice, m was 0.307, a difference of almost 50 percent), but did not increase the supply of words.

When the production of pleasant words was compared with the production of unpleasant words (Bousfield, *1944*), the difference was in the supply of words available rather than the rate of approach to this limit. A group of 71 students listed pleasant items and another group of 80 listed unpleasant items. (For the pleasant words c was 79.2 and for the unpleasant words 50.0, a difference of over 50 percent in the upper limit.) This effect of pleasantness and unpleasantness upon free recall has been studied for many years, but much of the confusion is cleared up by this kind of analysis. People know more pleasant than unpleasant words—or at least their limit of recall under these conditions is higher—though the rate of depletion of the supply is about the same for both.

Evidently the Bousfield and Sedgewick equation describes the continous production of words fairly well and variations in the two constants of the equation are psychologically meaningful. Further research

on this topic might profit by considering these two constants as well as the conventional raw score. One attempt to use the two analytical scores in correlational analysis will be described later in this chapter.

The constant, *c*, representing the supply of words is of course only estimated by extrapolation from a curve of perhaps eight data points obtained in 16 minutes. For something like naming automobiles a good fraction of the curve is obtained in 16 minutes, however, so probably little error is introduced in this way. More important is the fact that the supply constant is not a total supply in any absolute sense but is rather an effective or currently available supply. It is the supply that the subject's preparation makes available. Different conditions, especially different instructions, might give evidence of a larger or smaller supply. Study of lists of names of automobiles, as in the experiment mentioned above, does not increase the supply of automobile names, but redefinition of automobile or a set to name anything remotely resembling an automobile probably would increase the effective supply.

The constant *m*, representing the rate of depletion of the supply, is not a simple rate constant. The conventional raw score is a rate score since it measures the number of words produced in a standard time. The constant in the Bousfield and Sedgwick equation represents the rate at which the supply is being exhausted percentagewise. It remains constant, of course, throughout the production period, assuming that the curve fits the equation, while the simple number of words produced per unit of time decreases as production continues. The effects of frustration and of mood changes, mentioned above, probably are effects on the rate constant rather than on the supply constant. The supply of words is a result of preparation and is influenced by such factors as education, recent practice, and the immediate set for production, while the rate constant is influenced by the current level of functioning. When special practice is given just before production, the shape of the production curve may be changed slightly (Bousfield, Sedgewick, and Cohen, *1952*).

THE PROGRESS OF PRODUCTION

Marbe's law (see page 179), which states that, in single-word association, response time is shorter for common words than for rare words, could be extended to a hypothesis about continuous production of words. Words in common use will be produced earlier than less

common words. Bousfield and Barclay (*1950*) derive the same hypothesis from modern learning theory, specifically by extending Hull's principle of the habit-family hierarchy (Hull, *1934*). Words in most common use are reinforced most frequently and acquire strongest habit strength. The word with the strongest habit strength under the conditions of the experiment will be evoked first, the word with the next strongest habit strength will be evoked next, and so on. In general, for a group of subjects of the same cultural background, the order of occurrence of the words varies with the frequency of occurrence of these words.

Bousfield and Barclay tested this hypothesis by asking large samples of college students to name birds, carpenter's tools, and celestial bodies. When the order of occurrence of the words was plotted over their frequency of occurrence, clear relationships, supporting the hypothesis, were evident for all three kinds of words.

CLUSTERING. The overall set to name words is adopted by the subject in response to the experimenter's instructions. But even in this relatively simple assignment we can see signs of more specific sets, which change while the overall set to name words remains constant. Change of the specific set can be inferred from the relation between the response word and the stimulus word (see page 148). Clusters of responses may be seen in the following record, obtained by Bousfield and Sedgewick (*1944*) from one subject's production of names of U.S. cities. The subject was a student at the University of Connecticut. The dashes mark the two-minute intervals.

Hartford, Bridgeport, New York, Augusta, Boston, Buffalo, Schenectady, Montpelier, Concord, Providence, Trenton, Richmond, Waterbury, Nashville, Chicago, Atlanta, San Francisco, Los Angeles, Sacramento, Raleigh, Cheyenne, Austin, New Orleans, Lincoln,—Jefferson City, Kansas City, Topeka, Harrisburg, Columbus, Ann Arbor, New Haven, Denver, Phoenix, Miami, St. Petersburg, Salt Lake City, Lansing, Indianapolis, Duluth, Detroit, Columbia, Des Moines—Seattle, Salem, Washington, Bayonne, Carson City—Tuscumbia, Mobile, Stamford, Boston, Worcester, Rochester, Manchester, Cambridge, Burlington, Laconia—Syracuse, Middletown, Montgomery—Hollywood, Baltimore, New London, Keene—Charleston, Troy, Yorktown, Tampa, Poughkeepsie.

The persistence of the general set to name cities may be inferred from the fact that the subject did name cities for sixteen minutes. But it is obvious from the record that the cities are not drawn at random from

the subject's supply of cities. Apparently his first attack was to name nearby cities, for the first eleven are all fairly close to Connecticut. A cluster of three West Coast cities, San Francisco, Los Angeles, Sacramento, coming later, suggests another line of attack. The cluster, Seattle, Salem, Washington, suggests mediated generalization, by way of the state of Washington. Worcester, Rochester, and Manchester could be a cluster due to temporary domination by the ending, *ester*.

Words and ideas occur in clusters not only because of the set but also because of previous associations. This can be demonstrated by representing long lists of words, too long to be memorized, then asking for recall of words from these lists. Words in the same category, e.g., vegetables, are likely to be recalled in clusters in proportion to number of presentations (Bousfield and Cohen, *1952*). In recall under these conditions words of high frequency of usage, according to the Thorndike-Lorge word counts, cluster more than words of low frequency (Bousfield and Cohen, *1952a*). Likewise words that are frequently associated, according to the Kent-Rosanoff tables, will occur in two-word clusters (Jenkins and Russell, *1952*). The longer clusters illustrated in the above lists of cities are organized partly by past associations and partly by the approach to the problem at hand, that is, by the preparatory set.

The Function of the Set. It is commonly supposed that the set operates as a limit, keeping production within defined boundaries—defined in this case by the instructions from the experimenter as understood by the subject. But this would imply a skewed or U-shaped distribution of responses, with a surplus of erroneous or questionable responses concentrated near the limits. Most production records, however, show relatively few responses that do not conform to the instructions, so few that they are usually ignored in the scoring. It is more likely that in adopting a set the subject defines the goal response, e.g., the wanted word, precisely or loosely, correctly or erroneously, and that this search model acts as a center, facilitating production of responses close to this center and inhibiting deviate responses. If this is true, a distribution of responses, when plotted on a dimension of similarity to the goal response, as defined by the subject, would be unimodal.

The difficulty with this attempt to describe the distribution of responses produced under the control of a set is the lack of a scale or baseline of similarity to the goal response. It will be profitable for a

moment to consider a hypothetical experiment that would produce measurable responses. Suppose we assemble a large supply of wooden rods, like pencils, ranging in length from 100 mm. to 200 mm. by steps of one mm. We will place them in a basket, then ask someone to pick out all the rods that are 153 mm. long. Suppose he picks out 100 rods. Now if we measure each of these rods, we may expect that the distribution of lengths will be approximately normal. Any unimodal distribution will be evidence that the set operates, not as a limiting factor, but as a centering or facilitating factor. The mean will be the subject's concept of 153 mm., the goal response that he is set to execute. Both constant and variable errors could be computed. They would have the same meaning as in the psychophysical method of average error.

If the words that are produced to fit certain instructions could be scaled in the same way and a distribution plotted, some information on the operation of the set would be available. But if the definition of the goal operates as a centering point, deviations could occur along many semantic dimensions of similarity. Underwood and Hughes (see page 180) attempted to scale similarity of responses in an experiment on recall of paired associates and found the scaling rather difficult. To the extent that it was successful, however, there were many more responses judged in the highest degree of similarity than were judged in either of two lesser degrees of similarity. Counting the correct responses, which were most common, as maximum similarity, we would have a sharply peaked distribution on a baseline of similarity to the correct response. As far as it goes, this experiment adds further evidence that the set operates to center responses around a preferred response, deviations from which are less frequent.

This research and speculation on the production of simple responses like words have a certain unreal character. What has all this to do with solving genuine problems? The big difference, of course, is that in continuous production the words are being produced one after another merely to fit certain instructions, and that is the end of it. In thinking directed toward the solution of a verbal problem, the words are not just written down and forgotten. When a word that might be the solution to a problem is produced, the production process is halted, and the utility of the word is tested by an additional judgment process.

Let us return for a moment to our hypothetical experiment on selecting rods of a given length. Suppose we modify the instructions so that the task becomes a problem to be solved. We will use the same collec-

tion of rods of many lengths. We show our subject a broken piece of apparatus: a platform is uneven; it could be repaired with a leg of just the right length to level up the platform. Before we let our subject see the basket of rods, we might ask him how he proposes to repair the apparatus. He may disclose his preparation by saying, "I am going to look for a stick about so long to support the platform." If he does not talk about his preparation, we may still be able to infer it from his overt behavior.

If he now sees the basket of rods, he will probably select one and try it out. If he judges it suitable, that is the final act of judgment that concludes the problem-solving enterprise. Suppose we prevent him from trying out the rod he chooses. We will tell him to select now all the rods that may be useful and try them out later. By doing so we have experimentally magnified the production process, holding the preparation process constant and postponing the judgment process. This is the rationale for using the continuous production of words and other simple responses as exemplifying the control of problem-solving attempts by the preparatory set.

When one word is the solution to a problem, the earlier discussion of the order of production of words may be extended to a hypothesis about the difficulty of the problem. If the common words are produced first, the production of any particular word will take more time the less common that word is. Taking time required for solution as the measure of difficulty, the difficulty of a problem for which one word is the solution will be inversely related to the frequency of occurrence of that word in ordinary communication.

If, as our measure of difficulty of a problem, we take the number of people who fail it, some evidence on this hypothesis is available. The items of ordinary vocabulary tests are word problems of a simple sort, and the frequency of occurrence of each word is known. Our hypothesis asserts that if a sample of such test items is studied, the percentage of people who pass each item will be correlated with the frequency of occurrence of the word in ordinary communication. Two such correlations computed by Kirkpatrick and Cureton (*1949*) come out as .47 and .56.

Another difference between verbal problem solving and mere production of words is that in the former the preparation may not be given explicitly by the instructions. The subject may have to construct his own preparation and reconstruct it as he goes along. The produc-

tion process may appear functionally more significant if we ask our subjects to produce preparatory sets, schemas, definitions of the problem, plans of action, frames of reference, and criteria of success in continuous succession without stopping to test any. It is helpful to study each thought process in isolation, hoping that the isolation does not distort the process. As evidence that the process of production is not greatly distorted by this isolation, we shall note in the section on individual differences at the end of this chapter that fluency scores do have some practical predictive value.

PRODUCTION OF SOLUTIONS TO SIMPLE PROBLEMS

The production of potential solutions to a problem is harder to analyze than the continuous production of words because the production is not so regular and because the set frequently changes. In fact, as problem solving proceeds, the production process includes the production of new sets as well as of simple responses. Therefore there is no clear-cut separation between preparation and production. Though we may discuss three thought processes in separate chapters, problem solving telescopes all three together and makes temporal separation impractical. But the efforts of many ingenious experimenters have been directed toward the psychology of problem solving, and it is possible with the aid of the principles outlined in previous chapters, and a few new ones, to describe the events tolerably well. Considering the complexity and subtlety of the subject matter, progress has been encouraging.

A convenient illustration of a task that bridges the gap between continuous production of words and problem solving is one used by Bousfield and Sedgewick (*1944*) in the research described on page 183. They asked 21 college students to make words out of the letters ONDTERH. From their results, averaged for all 21 subjects, the curve of Figure 7.2 was drawn. It follows the same law as the other curves shown earlier and obviously portrays an approach to an upper limit at a decreasing rate.

AN ILLUSTRATIVE PROTOCOL. In order to illuminate the details of production of words in this simple case, one subject was asked to perform this task without hurrying and to comment on his method as he went along. He was asked to write and talk as much as possible. The

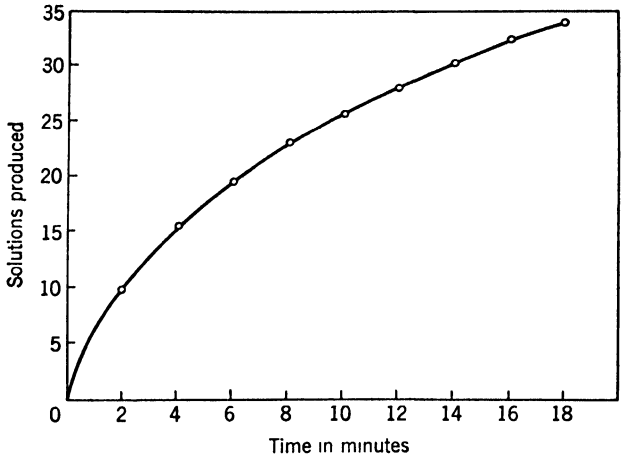


FIGURE 7.2. Course of production of solutions to an anagram problem: Make words out of ONDTERH. Average data for 21 subjects. Curve is drawn by eye. Data from Bousfield and Sedgewick (*1944*).

words written and the comments spoken are printed below, with breaks indicating, in the observer's opinion, a change of set.

1. tern
2. don
3. ton
4. her "Saw these at first glance." "I'll be systematic. Start with a common beginning letter."
5. don "Have it already."
6. den
7. doe
8. nod
9. net
10. hero "Just happened to think of this one."
11. hen
12. hot
13. her "Already have it."
14. thot "Probably not acceptable."
15. thore "Not a word."
16. or

17. thoer "No."
18. doer "See what happened?"
19. rod
20. rot
21. roe
22. ret "I think it is a verb."
23. rh "Nothing doing."
24. tre "No luck."
25. throe
26. trod
27. rho "No luck. This is getting dull."
28. dot "I don't know. Just saw it."
"I was wandering for a moment."
29. or "Have it."
30. oh
31. on
32. ode "Doing all right."
33. odor "No."
34. one (The subject pronounced it with a long *o*, wrote it slowly, then said "OK.")
35. ore "Do I have that?"
36. ert "No."
37. rote
38. eo "No."
39. edon "No."
40. done "Just luck."
41. hod "I am just staring at it now."

Results from other subjects have been similar but not always so revealing. Obviously this subject accepted the instructions from the experimenter and got set to produce words. The specifications of the response are clear-cut, and so is the stimulus material that he is to work with. The difference between this experiment and the usual word association experiment is that instead of having a simple stimulus word presented to him at a certain prepared-for moment, a relatively complex stimulus situation remains in view and he directs his efforts to different parts of it at different times.

The first four words were immediately apparent. Then the subject

surveyed the problem and adopted a more specific set—within the overall set to produce words. We could say that he constructed a search model, d_____, and looked for letters that would fill in the blanks.

He changes his specific set many times while retaining the general set to produce words from the letters ONDTERH. Not all subjects are as systematic as this subject was. And even in his case the responses were not completely controlled by the specific sets. Evidently the dominance of the specific set fluctuates for unknown reasons. Note responses 10, 28, and 41. Possible reasons for fluctuations of set were discussed in Chapter 3.

The dominance of the specific set is broken through also when the operations of the specific set disclose a stimulus pattern which, while it does not fit the requirements of that specific set, does fit the overall problem set. Note response 16, *or,* which does not begin with *th* but was disclosed when the unacceptable word *thore* was written down. Other examples are 18 and 40, apparently suggested by the preceding unacceptable words. Duncker (*1945*) uses the phrase "suggestion from below" for this phenomenon.

But a stimulus pattern at hand is not enough by itself to lead to a solution. The subject said the word *not* several times and the word *do* once, but he did not use them as problem responses, presumably because his overall set was directed toward the letters in ONDTERH, not toward his own conversation nor that of the experimenter. Or one might prefer to say that these two words were embedded in the pattern of his conversation, and he was not set toward analysis of this pattern.

SHIFTS OF ACTIVITY

Though the solutions to some problems may come from a straightforward and continuous production of responses, followed by evaluation of these responses, production of solutions to most problems is likely to be more variable. After producing a few responses of one kind, under the guidance of one set, the subject shifts set and tries something different. Hence many experimenters have found it necessary to analyze problem-solving activities by sequences, blocks, groups, or patterns of responses. When the unit of analysis is a complex molar one, extending over a period of time, the response pattern may be called an approach, attempt, attack, plan of action, or hypothesis; and variations in these may be more important than variations in the more molecular responses.

A shift in the set may easily be observed when the activity changes in kind, as when one who has just produced a block of bird names now produces a block of mammal names, or when the thinker who has been manipulating a puzzle with his hands reaches for a screwdriver, or when a lost automobilist who has been trying one road after another stops at a service station for advice.

As emphasized in the preceding chapter, preparation for responding includes an analysis of the problem situation. A certain stimulus or stimulus dimension, or stimulus configuration, is abstracted from the total environment, and the thinker gets set to respond to this feature of the environment. Under some conditions the shift of set may be chiefly on the perceptual side, a shift in what is responded to rather than in the response itself. What is observed by the experimenter is a change in the thinker's actions, to be sure, but a change in the perceptual side of the thinker's set may be inferred by procedures to be described below.

In these experiments the general set to observe and manipulate the problem material and to attain the requested solution is established by the experimenter's instructions, but the specific sets are task-induced (see page 70). That is, the responses may all be classified as problem-solving attempts, but some may be classified in more detail as one kind of attempt and others as another kind. Just as the production of responses of one kind, e.g., words, leads to the concept of fluency of production, variation of blocks or patterns of responses leads to the concept of flexibility of the set to perceive and respond.

Insight Versus Trial and Error. These two concepts, the history of which was reviewed briefly in the first chapter, have frequently been opposed to each other, and enthusiastic attempts to support one concept or the other with empirical data have added new techniques and new knowledge to our science. But no serious observer could ever have believed that problems are solved by either one without the other. If insight means understanding relationships within the problem, then those who have complete insight have no problem. And if trial and error means that the subject enters the problem situation blindly and responds at random, giving every aspect of the problem situation equal weight, it is a caricature of behavior. It must be admitted, if these terms are used, that insight and trial and error are complementary principles, both necessary for full-bodied description of problem solving; that insight into problematic situations changes, and usually improves, with repeated tries; and that even partial insights, at various

levels of complexity, limit the variation of the tries, usually reducing the errors.

Now that the old controversy has lost its sharp outlines, the present task is to describe the solution of problems as completely as possible, using the best concepts available, whether they coincide with the old ones or not. But the new concepts grew out of the same phenomena that gave birth to the old. There may be some value in bringing the old controversy up to date in terms of the concepts set forth in the present discussion.

As pointed out in the preceding chapter, the thoughtful subject prepares for a problem by organizing the data obtained from verbal instructions, the problem situation, and other sources. The better the organization of the data, the better the insight in the sense that production of responses is thereby guided, is less blind. But a strongly organized preparation for response may be unrewarding. Hence, even in the continuous production of words and certainly in complex problem-solving activity, the subject frequently reorganizes his preparation for response. He does so by resurveying the situation, breaking up the first pattern of the data, subtracting some aspects of the situation, adding some others, and restructuring the whole field. It seems likely that the laws of perception operate in organizing and reorganizing the data, especially if the problem is a spatial one, and probably similar laws operate when the problem is abstract.

These new insights or organizations of the data are not independent of practice; they result often from manipulation and exploration under the guidance of a previous preparation. This is what gives problem solving its trial-and-error character. Possible solutions are produced and tried out; then, if these are unrewarding, new ways of producing solutions are produced and tried out.

Since this production process is always, in any animal worth its salt, guided by a preparatory set, it is really the production of new sets, particularly task-induced sets, that challenges the ingenuity of psychological research. We must be able to record not only the production of responses but of sequences or patterns of responses, if necessary at a high level of complexity. It is for this reason that the shifting of set and the measurement of flexibility in shifting have come in for considerable attention from experimental psychology in recent years.

Recording Shifts of Set. Several procedures have been used to study the shifting of set in problem solving. They all involve the

establishment of a set; then, by a change in conditions, that set becomes ineffective, and another must be established. Shifting set when the conditions of the problem require it, i.e., when the existing set is unrewarding, is called flexibility. Failure to do so goes by such terms as perseveration, stereotypy, fixation, and rigidity.

Some of the early experiments on flexibility used the Yerkes multiple-choice apparatus or a modification of it suitable for human subjects. A number of doors or levers or boxes are exposed to the subject, who is to respond to one of them. The correct response is changed from trial to trial according to a system to be discovered by the subject. Repeated unrewarded choice of the same alternative is considered sterotypy or lack of flexibility.

In many serial discrimination problems only two alternatives are available, right and left, and the stimulus cards on right and left, which vary in several respects, are changed after each response. The stimuli may vary, for example, in respect to size, shape, and color. The subject, human or infrahuman, usually varies his responses until he discovers which stimulus aspect is rewarded. When the subject responds to the same stimulus aspect on every trial, in spite of irrelevant variations, he may be said to be set for that stimulus aspect.

Suppose, for example, that a rat on six trials in a discrimination problem turns L–R–R–L–R–L. Spatially his response has changed four times. But if some aspect of the stimulus pattern has been placed on the left, then on the right; then R–L–R–L, the rat's behavior, has, at a higher level, not changed at all, but has been directly related to that particular aspect. Thus we may infer his set. If, even though he is not rewarded, he remains set for that stimulus aspect, we may say that he is rigid or fixated. If his responses, when tabulated and analyzed, give evidence of shifting to another stimulus aspect, we may say that he is adaptable or flexible.

Flexibility at a more complex level is studied by arranging problems, e.g., arithmetic problems, word problems, which can easily and perhaps only be solved by one plan of action; then without comment introducing apparently similar problems which cannot be solved by that plan.

Some problems put the emphasis on flexibility in organizing and reorganizing patterns. The initial formulation of the problem is unrewarding, so the thinker must shift to a new organization of the

problem situation. Problems of this sort can be considered more conveniently in the next chapter.

CONDITIONS OF FLEXIBILITY

Robinson (*1940*) used a series of anagram games to study stereotypy. The object of each game was to make one six-letter word out of the jumbled letters on a stimulus card. All subjects began with the same card of letters, but one group of subjects had only that card, another could shift to an alternative card, and another group had six alternatives. The results showed that when the subjects had six alternative ways of reaching their goal, they shifted from one to another often, spending relatively little time on each card. When they had only one alternative, they persisted about four times as long.

EFFECTS OF PRACTICE. As stated in a previous chapter, a set may be thoroughly learned in one situation and reinstated later in a similar situation. Furthermore, a set adopted early in the performance of a task may persist and influence performance in the later part of the task. Such task-induced sets are of special importance for problem solving.

Luchins (*1942*) has demonstrated how a set that is useful in solving the first of a series of arithmetic problems may become a handicap in solving later problems. The problems all involved measuring required amounts of water, given two or three jars of stated volume. An unlimited supply of water was assumed. The details of the series of problems are given in the table below, taken from Luchin's report.

Problem	Given the Following Empty Jars as Measures			Obtain the Required Amount of Water
1.	29	3		20
2. E_1	21	127	3	100
3. E_2	14	163	25	99
4. E_3	18	43	10	5
5. E_4	9	42	6	21
6. E_5	20	59	4	31
7. C_1	23	49	3	20
8. C_2	15	39	3	18
9.	28	76	3	25
10. C_3	18	48	4	22
11. C_4	14	36	8	6

The first problem is introductory. The problem is put on the blackboard with the explanation that by filling the 29-quart jar, and pouring

from that into the 3-quart jar three times, 20 quarts will remain. The procedure may be written: $A - 3B$. The next five problems are intended to establish and mechanize the set or *Einstellung*. The procedure of the second problem, $B - A - 2C$, is explained, then the subjects are told to work the remainder by themselves. Hence the set is initially suggested by the instructions, then reinforced by its usefulness in solving the next few problems. The experimenter allows two and a half minutes for each problem.

Problems C_1 and C_2 are critical problems which test the persistence of the set. They can be solved by the usual procedure, $B - A - 2C$, but they can be solved much more directly in other ways: $A - C$ for the first one, and $A + C$ for the second one. Use of the more cumbersome method is considered evidence that the set or *Einstellung* established in the first few problems has persisted and interfered with the adoption of a simpler procedure.

Problem 9 is a particularly crucial problem because, unlike C_1 and C_2, it cannot be solved by the $B - A - 2C$ procedure. If anyone should begin with this problem, it would be very easy, as easy as the introductory problem. Will the set engendered by the first few problems be so strong that anyone will fail Problem 9?

The last two problems can be solved by either procedure, thus furnishing a further test of the persistence of the set.

Luchins tried out this series of volume-measurement problems with many groups of subjects, children and adults, varying the conditions slightly for some groups. Most subjects, 55 percent and more in various groups, solved C_1 and C_2 by the tedious $B - A - 2C$ procedure. Control groups that began with C_1 and C_2 did not use this method at all. Furthermore the majority of subjects failed to solve the crucial problem, No. 9, in two and a half minutes, though there were very few failures in the control groups. Attempting Problem 9 must have been helpful in breaking up the set for the tedious procedure because the college and adult groups used the direct procedure more often on C_3 and C_4 than on C_1 and C_2.

Luchins gave some groups special instructions. Before any problems were presented, these groups were told that, after completing Problem 6, they should write on their papers these words: *Don't be blind.* In general, in the college and adult groups at least, these instructions reduced the use of the indirect procedure on Problems C_1, C_2, C_3, and C_4 and also reduced failures on Problem 9.

In later experiments Luchins and Luchins (*1950*) tried several additional ways of reducing the mechanical effects of the set. They limited the supply of fluid by introducing the problem as one of measuring out required amounts of milk. In another experiment they added a fourth jar. And they made the problem more concrete by testing the subjects individually with actual jars appropriately labeled and a supply of water from a faucet (also Tresselt and Leeds, *1953a*). In general the effects of these variations in cutting down the influence of the set were disappointingly small. Some of these variations seemed to have different effects on susceptibility to set than on recovery from set.

In series of problems like these the set is acquired and strengthened by its usefulness in the first few problems. The set may be considered a higher-order habit, learned with a small but experimentally manipulable amount of practice. Luchins increased the amount of practice with the first procedure and found that this strengthened the set and increased the number of failures on the critical problems. This increase in the strength of the set with an increase in the amount of practice has been demonstrated also for the volume-measurement problems by Tresselt and Leeds (*1953b*), for other short arithmetic problems by Youtz (*1948*), and for anagram problems by Maltzman and Morrisett (*1952*).

Massing the practice with the first procedure—no time between problems—strengthens the set more than distributed practice—three minutes between problems (Kendler, Greenberg, and Richman, *1952*). The set does not decay quickly, as one might expect. When a delay of one day is interposed between the practice with the first procedure and the critical problems, the number of failures is not reduced but slightly increased (Tresselt and Leeds, *1953b*). The strength of the set is further increased slightly by delays of two and seven days. These two experiments, on distributed practice and delay of reinstatement, show that the development of the set does not follow exactly the same laws as the development of simple habits. As noted in Chapter 3, practice may be one important factor in the development of a set, and in these cases the set may be called a habit. But from the evidence at hand it should be qualified as a higher-order habit.

The effect of the set is strengthened also by compounding it. Maltzman and Morrisett (*1953*) found that practice in solving anagram problems by finding nature words will persist and influence the solution of later anagram problems, and so will practice in rearranging the

letters of the anagram in a specified order. Practice with a compound set, i.e., with anagrams that make nature words when they are rearranged in a specified order, produces more set solutions than practice with either simple set alone.

From these experiments it is clear that practice does not always make people perfect; it sometimes makes them rigid. Other experiments show that one cure for these deleterious practice effects is more practice at a higher level. Luchins (*1942*), for example, noted that increasing the number of critical problems facilitated recovery from the set. Grant and Berg (*1948*) had their subjects, doing a serial discrimination problem, shift several times so that they got practice in shifting. Responding on the basis of color was called "right" at first; then, without comment, number classifications were called "right" and the others "wrong." In a similar way, the basis of correct classification was shifted to form, back to number, then to color, and finally to form again. Seven groups of college students did these problems, one of which was allowed three correct trials with each classification before the shift to the next, another was allowed four, another five, and so on to ten confirming or reinforcing trials. It turned out that the groups with larger number of preshift reinforcing trials made fewer errors, both perseverative and nonperseverative, after the shifts. Evidently the subjects who were allowed the additional reinforcing trials had a chance to master each problem thoroughly and thus to become more sensitive to a change in the pattern of reinforcement. Different schedules of reinforcement have different effects on the ability to shift (Buss, *1952;* Schroder and Rotter, *1952*). In general the more varied patterns of training favor the development of flexible higher-order habits.

The development of flexible higher-order habits is not limited to college students. Harlow (*1949*) was able to train apes and children, as well as adults, on a serial discrimination reversal problem so that shifting to a new problem was very rapid.

THE EFFECTS OF HASTE. In some of Luchins' experiments he speeded up his subjects by shortening the time allotted for working each problem from two and a half minutes to one minute and less and by instructions to hurry. This speeding up decreased the flexibility of his subjects. Other groups were told not to hurry, to take their time—and some used as much as six to ten minutes on one problem—but this did not increase flexibility.

Rokeach (*1950*) checked on the influence of haste in working these

volume problems, not by allowing the subjects more working time, but by forcing them to delay before starting work. He presented the problems on the board, but told the subjects not to begin until they were permitted. Four groups of college students were used and delays of 10, 30, 30, and 60 seconds. Significant differences due to the differences in exposure time were found. Apparently the enforced delay interrupts the set and allows competition from other sets.

Effects of Emotion and Frustration. Hamilton's (*1916*) pioneer experiment with several species of mammals demonstrated that emotional excitement reduces flexibility of problem-solving attempts. A frustrated animal will often persist in an unrewarding pattern of behavior and thus fail to solve a problem that it could solve when not excited (Hamilton and Krechevsky, *1933;* Everall, *1935*). Birch (*1945*) found that, when chimpanzees attempt difficult problems requiring insight, intermediate degrees of motivation are the most effective. Low degrees of motivation permit distraction into irrelevant activities, and high degrees of motivation (36 and 48 hours of food deprivation) produce frustration at failure and inability to shift from an unsuccessful attack to a new one. A series of experiments on the rat by Maier (*1949*) and his students has demonstrated in some detail how frustration produces fixation of unrewarding response patterns.

One of the most thorough experiments on the effect of emotion on problem solving is that of Patrick (*1934*), described in Chapter 2. Prominent in the results under emotional conditions is the increase in stereotyped and perseverative reactions (see Figure 2.3).

Maier's technique for studying the effects of frustration on rats in a serial discrimination problem has been applied to human subjects by Marquart (*1948*). The stimuli are presented on two small doors, as in the jumping apparatus used with rats, but each door is opened by a metal handle that can deliver an electric shock to the subject's hand. The stimuli are of various shapes and sizes, arranged in various combinations. The subject is told to find "a pattern running throughout the situation." A problem without a solution is presented first, the subject being arbitrarily punished regardless of his choices. Then a simple position problem is presented, the correct stimulus card always being on the left.

The most valuable results were obtained with a group of 42 college students. They were arbitrarily punished 75 percent of the time for 50 trials; then, without comment, the door on the left was made the cor-

rect one, i.e., opening the left door was not followed by shock. Learning a position habit, such as always turning the handle on the left, would of course be very easy if not preceded by the frustration of 50 arbitrarily punished trials. As it turned out, 11 subjects required more than 120 trials to master this simple problem and two failed entirely. The mean number of trials required for mastery was 69 (see Figure 7.3).

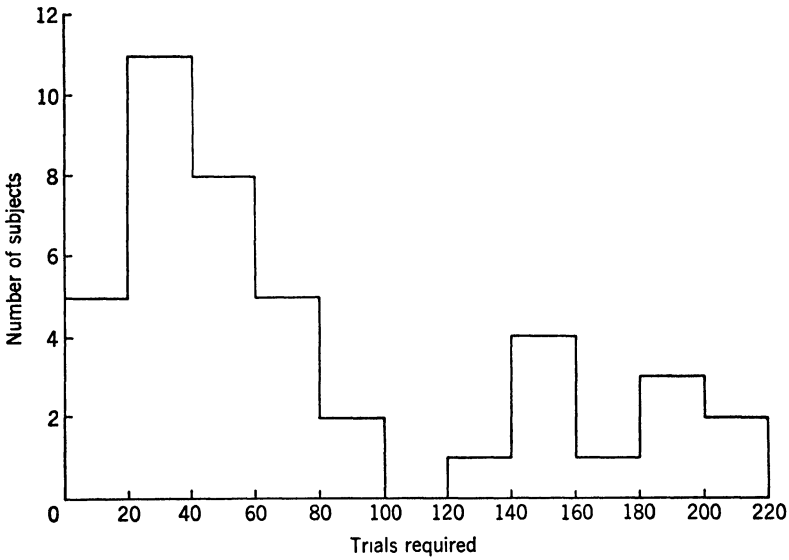


FIGURE 7.3. Bimodality in problem-solving scores produced by frustration. Scores from 42 subjects working a serial discrimination problem after arbitrary punishment. From Marquart (*1948*).

To analyze the role of flexibility in this experiment Marquart computed an index of stereotypy, namely, the number of repetitions of choice of stimulus cards after the shift from the no-solution problem to the position problem. According to this index the slow learners were much more stereotyped, or less flexible, than the normal learners.

From what has gone before we would expect that flexibility in shifting set on the Luchins water-jar problems would be reduced by emotion. Cowen (*1952*) put some of his subjects under mild stress by giving them a practically insoluble puzzle, and put some under strong stress by telling them of pathological signs in their projective test

records. Those under strong stress gave 5.1 rigid solutions to the jar problems, and those under mild stress gave 2.5, as compared to 1.2 for the controls. The same effect might be expected on the Wisconsin Card Sorting Test (see page 209), but Ross, Rupel, and Grant (*1952*) found that heckling and distraction produced only small disturbances. Only electric shock lowered performance significantly.

If we speculate on the kind of attitude or mood that would be most free from stress and therefore most conducive to flexibility of thought, the answer seems to be play. The thinker who is just playing with words, colors, sounds, or ideas, with no great stake in the outcome and no fear of failure, has nothing to stop him from shifting. He can follow his flights of fancy wherever they lead him.

INDIVIDUAL DIFFERENCES IN PRODUCTION OF SIMPLE RESPONSES

The common complaint against the separation of experimental psychology and differential psychology does not apply to the topic of this chapter, for the production of simple responses and sequences of responses has been studied from both points of view. The available information on individual differences in production is valuable in itself and also furthers understanding of the general principles of production.

Differences between people in the production of single words, as in free association, are largely differences in preparation and thus were considered in the preceding chapter. Our principal topics now are fluency and flexibility (or its opposite, rigidity).

FLUENCY SCORES

The continuous production of words was used as a psychological test before this task was investigated experimentally. The early Stanford-Binet Test (Terman, *1916*) had an item at the 10-year level that asks the child: "Name all the words you can think of." In England a quarter-century ago Hargreaves (*1927*) gave a battery of tests calculated to get at the "faculty" of imagination and identified, by methods then available, a common factor in tests demanding a ready spontaneous flow of indiscriminate words, images, etc. This factor, which seemed to be different from *g* and perseveration, has become known as fluency of imagination, verbal fluency, or fluency of association.

In the United States fluency has usually been treated as an ability

factor. The word-naming item on the Stanford-Binet Test has an important correlation with general intelligence. The Chicago Tests of Primary Mental Abilities (Thurstone and Thurstone, *1943*) include one that requires naming words that begin with a given letter and one that requires naming four-letter words that begin with another given letter. The composite of these two is called W for word fluency.

Thurstone (*1938*) has demonstrated, in analyzing a battery which included several kinds of verbal tests, that W, word fluency, could be separated from V, verbal comprehension, as measured by vocabulary tests. The separation is not complete, however. In a sample of high-school students the correlation between scores for W and for V was .54 (Thurstone and Thurstone, *1943*). In this sample the primary ability, W, correlated .69 with the general factor.

Johnson and Reynolds (*1941*) approached this topic from a different angle and reached a similar conclusion. They were interested in problem solving and set up the hypothesis, based on experimental research, that success in solving problems would depend upon (1) the flow or production of responses and (2) the selection of responses appropriate to the problem. Confining the analysis to verbal problems, they constructed tests that ranged from pure flow of words to a high degree of selection. When these tests were given to college students, factor analysis of the correlations disclosed two independent factors as expected: production of words and selection of words. Surely these factors are similar to Thurstone's W and V, but the hypothesis that led to their identification describes somewhat more dynamically the nature of their joint contribution to problem solving. Only the production process is discussed in this chapter; the selection or judgment process will be discussed in later chapters.

Other factor analyses in this domain (Carroll, *1941;* Gibb, *1942;* Taylor, *1947;* Fruchter, *1948*) have subdivided verbal fluency in various ways. It seems likely that ideational fluency, which requires the production of words or short phrases in a meaningful way, can be separated from simple nonmeaningful word fluency. This new factor of ideational fluency may turn out to have practical importance in predicting success on various verbal assignments (see Chapter 11).

In England (Cattell, *1934;* Studman, *1935*) and more recently in America fluency has been studied as a personality trait. Cattell (*1946*), for example, summarizing many factorial studies of personality, lists fluency as a strong factor in the primary personality trait that he calls

"surgency." The essence of surgency "is a placid, unemotional, realistic cheerfulness, with talkativeness, geniality, enthusiasm, sociability, and a witty originality." Finding fluency and cheerfulness associated factorially is to be expected from the demonstrated effect of frustration on fluency (see page 202). The frustrated, depressed person cannot loosen up and produce words spontaneously for the same reason that a person walking a tightrope across Niagara Falls cannot dance spontaneously. It is the cheerful, secure person who is willing to let go. The manic is walking a tightrope also but he doesn't care. (See also Vernon, 1950.)

Although fluency tests are easy to administer and score, there is little evidence on the distribution of fluency in different samples of the population. In one of the few investigations of this problem, Cattell and Molteno (*1940*) compared monozygotic with dizygotic twins in respect to intratwin differences in fluency, and found slightly smaller differences in the former. Comparing these with differences between comparable individuals chosen at random, they conclude that the environment is about eight times as important as heredity in producing variations in fluency scores.

Use of Analytical Scores. The conventional fluency score, which may be called a raw score, is the number of words or other items produced in a limited time. The equation of Bousfield and Sedgewick describing the curve of production of words (see page 183) suggests two analytical or derived scores that may yield more information than the single raw score. In this equation, $N = C\ (1 - e^{-mt})$, C is a constant representing the upper limit or supply of words and m is another constant representing rate of depletion of the supply. C and m can be calculated for each subject, if the production curve is reasonably regular, and may be then treated as analytical scores.

Johnson, Johnson, and Mark (*1951*) gave two tests of fluency to 43 college students: "Write names of U.S. cities for 15 minutes." "Write animal names of two or more syllables for 15 minutes." Production curves were drawn for each subject for each test and from these the constants C and m were computed. The usual raw fluency score was also available for each test. The intercorrelations showed, first, that the usual fluency score was more closely related to C, the upper limit constant, than to m, the rate constant, in both of these 15-minute tests. When the fluency score is obtained from shorter tests, one may expect that it will be more dependent upon the rate constant. Second, the constants, C and m, were shown to be inversely correlated for each of the

two tests. The meaning of this unusual result is not clear, but it points to some kind of an interference effect. When many words are available, they are not emitted as fast, relatively, as when there are fewer. And third, the correlations suggest that the rate constant, *m*, is more general than the upper limit constant, *C*. At least on these two tests the correlation between *m* and *m* was higher than that between *C* and *C*. It is possible that the use of analytical scores of this sort would aid the interpretation of factorial studies of fluency and related tests.

FLEXIBILITY SCORES

Individual differences in ability to maintain a set were discussed in Chapter 3. General principles of shifting set were discussed earlier in the present chapter. We are now concerned with individual differences in ability to shift from one set to another during the solution of a problem. The flexibility of human beings in problem situations has often been contrasted with the stereotyped behavior of infrahuman animals. Attempts to get scores for individual differences in this flexibility have taken several forms.

PERSEVERATION AND ALTERNATION. The trait known as perseveration, long studied in England, appears at first glance to be related to lack of flexibility in shifting set. It is best tested, according to Cattell (*1946*), by tests that require a reversal of established habits, e.g., writing letters or numbers from right to left. Rate of production of these new responses is compared to rate of production of the old ones. To the extent that the new performances are slow relative to the old, the subject is said to stand high on perseveration or "disposition rigidity." It is a narrow factor, clearly identified only in motor performances.

Kleemeier and Dudek (*1950*) have recently analyzed alternation tests factorially. Their subjects had, for example, to add, to subtract, then to do both alternately. The test battery was built around the numerical, verbal, and perceptual speed factors, and after rotation these well-known factors appeared, but no flexibility factor was necessary. A study by Guetzkow, to be described below, failed to show any relation between perseveration of this kind and flexibility in solving problems.

If these tests are tests of flexibility in shifting set, they must involve sets of a very low order of complexity. At any rate it is not likely that this simple flexibility or perseveration is related in any important way to flexibility in producing solutions to problems.

Flexibility in Grouping and Sorting Perceived Objects. The tests that are most pertinent to flexibility of thinking require adopting a set and later shifting to another set. The question is not one of speed of performing the new task, as in alternation and perseveration tests, but rather whether the new task can be performed at all, and in what way. A number of grouping or sorting tests have been devised for use in testing neuropsychiatric patients, several of which are of interest at this point. (See Vigotsky, *1934;* Hanfmann and Kasanin, *1937,* Bolles, Rosen, and Landis, *1938;* Weigl, *1941;* Hunt and Cofer, *1944;* Rapaport, *1945.*)

In general the procedure of sorting tests is to spread out before the patient a number of blocks or other objects with instructions to "put the ones together that belong together." The patient may group on the basis of color, putting the green objects in one pile, the red in another, and so on. He may next be asked to group the objects in some other way, or the problem may be such that his first grouping is obviously inadequate. Thus he has to shift set and group according to size, shape, use, or some other common characteristic. Usually he is asked to describe the basis of his grouping.

Most patients with cerebral lesions do not perform on these tests as well as normal subjects, and many schizophrenics are somewhat handicapped (Hunt and Cofer, *1944*). But there is no clear agreement as to the nature of the psychopathology. Some clinicians refer to the defect disclosed by sorting tests as an inability to abstract, while others refer to it as an inability to shift set. The matter requires a little discussion.

In order to put things together that belong together, one must be able to abstract a common characteristic, e.g., color, from a variety of other observable characteristics. One must maintain a set to perceive that stimulus aspect and ignore others, e.g., size and shape, long enough to complete the grouping. Failure can be due to either inability to abstract or inability to maintain the set, or both. It follows from the definition of set in Chapter 3 that maintaining a set under changing conditions demands that the subject abstract from the complex changing surroundings the stimulus or stimulus aspect to which he is responding. In reverse, abstracting, generalizing, noting similarities and differences in a collection of objects, always demands maintenance of a set toward the common characteristic in spite of variations in location and associated characteristics. Hence abstracting and maintaining a set must always go together.

But tasks can be arranged in which the abstracting or generalizing ability is crucial, and others can be arranged in which the ability to maintain and shift set is crucial. For example, if the patient can group objects according to one characteristic, e.g., size, but cannot shift when requested to another way of grouping them which by itself would not be more difficult, e.g., color, the defect must be an inability to shift set rather than an inability to abstract. Thus we see that the ability to abstract or generalize and the ability to shift set can be separated by the arrangement of the task and by the scoring procedure. Since empirical data on the relationships between the various tests of this type are not available, such separation for convenience of discussion can only be speculative. The next chapter, which deals with the formation of concepts and other abstract patterns, will therefore consider those tests that seem to emphasize abstraction, generalization, and concept formation, and tests that seem to emphasize flexibility in shifting set will be taken up now.

The Hanfmann-Kasanin Test (*1937*) may be taken as representative of one type of sorting test. In this test 22 blocks of various sizes, shapes, and colors are displayed to the subject on a table. The examiner states that there are four kinds of blocks, the names being concealed on the bottom. He turns one block over, showing the subject its name. The subject is instructed to find other blocks like this one, put them in one corner of the table, and put the other three kinds in the other three corners. The subject may begin to group the blocks on the basis of color, but then see that there are five colors. He may begin with shape, but there are more than four shapes. The examiner can correct the subject, if his grouping is wrong, by turning over two or more blocks in a group, showing that they have different names. An overall score may be obtained from the time required and the number of corrections given, but many examiners prefer to make qualitative observations. If, for example, a subject begins to group on the basis of color and cannot shift to grouping on some other basis, he may be said to lack flexibility. Rapaport (*1945*) suggests observing four aspects of thinking: flexibility, fluidity, persistence, and rigidity.

In the usual grouping test the objects to be grouped are spread out before the subject. The Wisconsin Card Sorting Test (Grant and Berg, *1948;* Berg, *1948*) uses a serial presentation. A large number of cards are to be sorted into four bins according to criteria to be discovered by the subject. He is told when his responses are correct, but after he has

discovered the criterion and is sorting correctly, the criterion is changed without comment, and he finds that he is no longer sorting correctly. He must shift to a new set. He is not told to shift, but his responses are not rewarded until he does shift. Errors due to perseveration or inability to shift may be objectively separated from errors due to other difficulties. Hence it appears that the use of this test and comparison of the scores with scores from other tests would help to clarify some of the confusion in this area. Scores on the test have only low correlations with general intelligence, and they do (especially the perseverative responses) separate the brain-injured from the normals (Berg, unpublished) and schizophrenics from normals (Fey, *1951*).

The task used by Marquart (*1948*) in her study of the effect of frustration on flexibility was a two-choice serial discrimination problem similar to serial sorting tests in that the solution of the second task required shifting from the set induced by the first one (see page 202). A distribution of scores representing trials necessary for mastery of the second problem shows that the frustration divided the 42 subjects into two different groups (see Figure 7.4). Similar results are produced by a no-solution problem even without electric shock (Marquart and Arnold, *1952*). Such discontinuity is extremely rare in distributions of scores of significant performances. It is produced by an emotional variable of considerable importance for psychopathology. The procedure of introducing a sorting problem by a frustrating no-solution problem holds much promise for comparisons of normal and neuropsychiatric subjects.

Flexibility in Shifting Attack on a Problem. These sorting tests emphasize flexibility on the perceptual side of the set. The response, i.e., the sorting operation itself, is easy; the difficulties that determine the scores must be difficulties in shifting from a set to perceive one stimulus aspect to a set to perceive another. Some sorting tests, however, e.g., the Hanfmann-Kasanin, permit qualitative observations of the mode of attack. And some techniques that have been employed in the study of flexibility of thinking specifically emphasize shifting the attack on the problem, changing the pattern of action.

Several investigators, following the early work of the comparative psychologists, have used a measure of stereotypy, that is, some index of the number of kinds of attack on the problem. Guetzkow (*1951*), in a study to be described below, got an index of stereotypy or inclination toward habitual solutions by dividing the number of attempted solu-

tions to a problem by the number of different types of attempted solutions. Marquart used the number of repetitions of response in a two-choice problem situation.

Another technique for obtaining scores emphasizing individual differences in flexibility of problem-solving attacks may be illustrated by an examination of the series of volume-measurement problems (see page 198) used by Luchins (*1942*). It will be recalled that the experimenter's demonstration and the subject's experience with the first few problems establish a set to attack the problems by following the demonstrated procedure. The next few problems, called ambiguous problems, can be done by following the same procedure or by shifting to a new and shorter procedure. The number of problems done by the old cumbersome method may be taken as a score for susceptibility to the set. Then follows a problem that can only be solved by shifting to the new short procedure. Failure to solve this critical problem is an indication of rigidity or inability to recover from the set previously established. If several recovery problems are given, the rigidity score for any subject would be the number of such problems failed. It is possible, though the evidence is not complete, that susceptibility to set and inability to recover from set, as thus measured, are somewhat different aspects of flexibility (Guetzkow, *1951*). If this is true, one would expect that the ability to recover from set when the situation requires it would be an ability variable, whereas the tendency to shift when the situation is ambiguous would be more of a personality variable.

The general program of establishing a set, then requiring that it be altered, lends itself to other kinds of material. Rees and Israel (*1935*) and Luchins (*1942*) have used series of hidden-word problems. In the first few groups of letters, animal words are easily discovered by a demonstrated procedure; then in later problems other words, not animal words, which are more obvious, are overlooked by those who rigidly follow the practiced procedure. Luchins (*1942*) and Rokeach (*1948*) have used mazes and obtained scores for inability to shift from a practiced, circuitous pathway to a short one. Cowen, Wiener, and Hess (*1953*) devised an alphabet maze according to the same strategy.

Comparisons and Correlations. Any serious attempt to use these kinds of scores for investigation of individual differences in flexibility of problem solving immediately raises the question of reliability. The usual measures of internal consistency are not applicable because the separate problems make different contributions to the score according

to their positions in the series. It might be feasible to prepare two equivalent series of problems for calculation of repeat reliability. At present tests of flexibility are of unknown reliability, and the reliability question is important because all published correlations are low.

The next question is one of generality of flexibility. To date only Cowen, Wiener, and Hess (*1953*) have attacked this question directly, i.e., by correlating scores obtained from two tests of the same form but of different content. One was the Luchins series of arithmetic problems and the other was a parallel series of alphabet maze problems. Beginning with 59 college students and eliminating 12 for various reasons, these experimenters got two scores for each of the remaining 47 students. Both distributions of scores were U-shaped. The minimum estimate of the correlation between the two scores was .42. Hence a moderate degree of generality was demonstrated in this small sample of college students.

Several minor factor analyses have attempted to discover the relationships among tests in the general area of rigidity. Cattell and Tiner (*1949*), for example, factored the results from 100 college students who took 17 tests. The disposition rigidity or perseveration factor appeared and another rigidity factor, of more promise for problem solving, tentatively called ideational inertia. It is associated with low scores for reconstructing hidden words and inventing answers to riddles. In this and other factor analyses fluency scores have been found to have a low negative relation to rigidity factors.

Oliver and Ferguson (*1951*) included rigidity tests in a factorial study with 98 college students. Intercorrelations were low. The Luchins series of jar problems had its highest correlation (.354) with a number series test, its next highest (.262) with a figure analogies test, and next (.245) with the Gottschaldt figures. These tests and a few others identified a factor that was called reasoning. A weak factor, identified principally in tests that required reversal of habitual operations, was called rigidity. This is the low-level rigidity or perseveration factor, not the problem-solving rigidity of the Luchins jar problems.

Some of the difficulties involved in analyzing the various traits in the realm of flexibility and rigidity have been discussed by Cattell and Winder (*1952*). To identify a trait that can properly be called flexibility, it is necessary to eliminate success due to general intelligence and such group factors as verbal and numerical ability, as well as simple perseveration and fluency factors. At present there is no good evidence

for a general problem-solving flexibility or rigidity factor of any importance. The results of Cowen, Wiener, and Hess indicate, however, that more positive conclusions may be reached if several tests of flexibility in shifting set are constructed on the same plan.

Another, and much easier, approach to this area is to begin with a test assumed to measure flexibility in shifting set and find what variables are related to it. The Luchins series of jar problems has captured the attention of many psychologists and has frequently been taken as a measure of flexibility to be compared with scores from other tests.

Guetzkow (*1951*) compared scores on the Luchins series of jar problems with Maier's well known two-string problem (see page 256) adapted for group administration. The instructions for the two-string problem ask for many types of solution, three of which were assumed, on the basis of previous work, to be reproductive or habitual, while a fourth, the pendulum solution, was assumed to indicate productive thought. An index of stereotypy or inclination toward habitual solutions was obtained by dividing the number of solutions by the number of different types of solution. As to the question of generality of flexibility, Guetzkow's results, as far as they go, agree with the negative conclusions of most previous work. The subjects who clung to the habitual method on the jar problems were only slightly more stereotyped on the two-string problem than those who shifted their approach. Similarly, those who failed the recovery problem of the jar series offered scarcely less pendulum solutions to the two-string problem than those who solved it.

For further analysis of susceptibility to set and ability to overcome set, Guetzkow made up two pairs of extreme groups: 22 who were especially susceptible to set on both problems, for contrast with 25 nonsusceptibles; and 25 who solved the recovery problem of the jar series and also gave pendulum solutions to the two-string problem, for contrast with 47 who failed to do either. These extreme groups were then given tests which have been analyzed factorially, together with a 12-item reasoning test constructed of the kind of problems used by Duncker, Maier, and Wertheimer (see Chapter 8). In respect to susceptibility to set, the nonsusceptibles scored higher on the Thurstone flags, the Gottschaldt figures, and the 12-item reasoning tests; but lower on the American Council on Education Test (ACE), presumably because on the first three tests much shifting is required, while on the ACE continuance of a set is helpful. In respect to ability

to recover from set, the solvers did better than the failures on the Gottschaldt figures and on a test of finding concealed animals. These last two tests "demand freeing the desired figure from the configuration in which it occurs."

Other investigators have shown that scores on the Luchins volume-measurement or jar problems have little or no relation to arithmetic ability (Rokeach, *1948;* Solomon, *1951*), but a low positive correlation with general intelligence (Rokeach, *1948;* Cowen and Thompson, *1951*) and verbal abilities (Solomon, *1951*).

Terms like flexibility and rigidity can refer to personality as well as intellectual traits. From what is known about authoritarian and ethnocentric people, Rokeach (*1948*) suspected that people who score high on ethnocentrism would be more rigid in a problem-solving situation than those who score low. Therefore he compared 35 college students who scored above the median on an ethnocentrism test with 35 who scored below the median. The highly ethnocentric students solved 2.2 jar problems in a rigid way, on the average, while the less ethnocentric half solved only 1.4 problems rigidly. The difference was small but significant, and it has been confirmed by Solomon (*1951*). Apparently the relationship between rigidity and attitude is high enough to yield a significant difference between the means of extreme groups, but not high enough to yield a correlation coefficient of any importance (Goodstein, *1953*).

Cowen and Thompson (*1951*) compared 17 rigid and 17 flexible eighth-grade children, classified thus in respect to performance on the Luchins jar problems, as to scores on printed personality tests and Rorschach responses. The personality tests did not differentiate the two groups, but some of the Rorschach signs did. When expert judges rated the Rorschach records for overall adjustment and for rigidity, those who were rigid in solving the problems were rated low in overall adjustment and high in rigidity, though only the first difference was statistically significant. Rigidity on the Luchins jar problems correlates in the .40's with a variable called "willingness to admit that a response is a reasonable interpretation of a Rorschach card," and with fluency as measured by number of perceptual hypotheses given to briefly exposed objects (Ericksen and Eisenstein, *1953*). Low correlations with scores on a printed test of anxiety have also been reported (Maltzman, Fox, and Morrisett, *1953*).

Some unpublished data on the relation between problem-solving

rigidity and various personality variables have been kindly made available by M. Rokeach. Scores for 72 college students on the jar series, with ten critical problems, were correlated (point biserial *r*) with scores from the following personality tests:

Anxiety scale, abstracted by Welch from Minnesota Multiphasic Personality Inventory	.383
Rigidity scale, constructed by Gough and Sanford	.334
Dogmatism scale, constructed by Rokeach	.285
Authoritarianism, the F scale	.276

Another printed test intended to give scores for rigidity has been described by Wesley (*1953*). The items are such as:

There is usually one best way to solve most problems.
I prefer work that requires a great deal of attention to detail.
I find it easy to stick to a certain schedule once I have started on it.

When college students with high scores on this test were compared to students with low scores in respect to performance on a modification of the Wisconsin Card Sorting Test (see page 209), the high scorers made more errors, especially more perseverative errors.

These studies of individual differences in flexibility of changing set during problem solving have not gone far. Compared to the analysis of fluency scores, for example, the analysis of flexibility scores is just beginning. It is not yet clear to what extent susceptibility to set, as indicated by persistence in a circuitous attack when a more direct one is available, is related to inability to recover from set, as indicated by outright failure on a problem that can only be solved by changing set. Interpersonal relations with the experimenter have not been seriously studied, although the set is initiated by the experimenter's activities and in some respects the experiment resembles an experiment on prestige suggestion. (Some of the factors in the social situation that influence the subject's set for problem solving were listed in the preceding chapter.) Methodological difficulties of rigidity tests have been pointed out by Goodstein (*1953*) and by Levitt and Zelen (*1953*). The most general criticism is that the conventional techniques of test construction, refinement, and evaluation, usually applied to tests of aptitude and personality, have simply been ignored.

Most of the present evidence—the single exception is Cowen, Wiener,

and Hess—indicates that change-of-set scores are specific to the task rather than measures of a general flexibility factor. The existence of a broad flexibility factor, independent of other well-known factors that can be tested more conveniently, has not yet been proved. On the other hand, the particular change-of-set tasks that have been used, e.g., flexibility in arithmetic problem solving, do have some promising relations to certain intellectual variables, e.g., reasoning and finding concealed figures, and to certain personality variables.

REFERENCES

Baker, L. M., and Elliott, D. N., 1948. Controlled and free association-times with identical stimulus- and response-words, *Am. J. Psychol., 61*:535–539.

Berg, E. A., 1948. A simple objective technique for measuring flexibility in thinking, *J. Gen. Psychol., 39*:15–22.

Birch, H. G., 1945. The role of motivational factors in insightful problem solving, *J. Comp. Psychol., 38*:295–317.

Bolles, M. M., Rosen, G. P., and Landis, C., 1938. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia, *Psychiatric Quart., 12*:733–737.

Bousfield, W. A., 1944. An empirical study of the production of affectively toned items, *J. Gen. Psychol., 30*:205–215.

Bousfield, W. A., and Barclay, W. D., 1950. The relationship between order and frequency of occurrence of restricted associative responses, *J. Exper. Psychol., 40*:643–647.

Bousfield, W. A., and Cohen, B. H., 1952a. *The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates,* Technical Report No. 1 Office of Naval Research Contract Nonr —631 (00), University of Connecticut.

Bousfield, W. A., and Cohen, B. H., 1952b. *The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-ciates,* Technical Report No. 1, Office of Naval Research Contract Nonr —631 (00), University of Connecticut.

Bousfield, W. A., Sedgewick, C. H. W., and Cohen, B. H., 1952. *Certain temporal characteristics of the recall of verbal associates,* Technical Report No. 3, Office of Naval Research Contract Nonr—631 (00), University of Connecticut.

Bousfield, W. A., and Sedgewick, C. H. W., 1944. An analysis of sequences of restricted associative responses, *J. Gen. Psychol., 30*:149–165.

Buss, A. H., 1952. Some determinants of rigidity in discrimination-reversal learning, *J. Exper. Psychol., 44*:222–227.

Carroll, J. B., 1941. A factor analysis of verbal abilities, *Psychometrika, 6*:279–307.

Cattell, R. B., 1934. Temperament tests: 2. Tests, *Brit. J. Psychol., 24*:20–49.
Cattell, R. B., 1946. *Description and measurement of personality,* Yonkers, World Book Co.
Cattell, R. B., and Molteno, E. V., 1940. Contributions concerning mental inheritance: II. Temperament, *J. Genet. Psychol., 57*:31–47.
Cattell, R. B., and Tiner, L. G., 1949. The varieties of structural rigidity, *J. Personality, 17*:321–341.
Cattell, R. B., and Winder, A. E., 1952. Structural rigidity in relation to learning theory and clinical psychology, *Psychol. Rev., 59*:23–39.
Cowen, E. L., 1952. The influence of varying degrees of psychological stress on problem-solving rigidity, *J. Abnorm. & Social Psychol., 47*:43–63.
Cowen, E. L., and Thompson, G. G., 1951. Problem solving rigidity and personality structure, *J. Abnorm. & Social Psychol., 46*:165–176.
Cowen, E. L., Wiener, M., and Hess, J., 1953. Generalization of problem-solving rigidity, *J. cons. Psychol., 17*:100–103.
Dewey, J., 1910. *How we think,* Boston, D. C. Heath & Co.
Duncker, K., 1945. *On problem-solving,* Psychol. Monogr., 58. No. 270.
Ericksen, C. W., and Eisenstein, D., 1953. Personality rigidity and the Rorschach, *J. Personality, 21*:386–391.
Everall, E. E., 1935. Perseveration in the rat, *J. Comp. Psychol., 19*:343–368.
Fey, E. T., 1951. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting test, *J. Cons. Psychol., 15*:311–319.
Fruchter, B., 1948. The nature of verbal fluency, *Educ. Psychol. Meas., 8*:33–47.
Gibb, C. A., 1942. Personality traits by factorial analysis, II, *Australasian J. Psychol. and Philos.,* Sept., 27.
Goodstein, L. D., 1953. Intellectual rigidity and social attitudes, *J. Abnorm. & Social Psychol., 48*:345–353.
Grant, D. A., and Berg, E. A., 1948. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem, *J. Exper. Psychol., 38*:404–411.
Guetzkow, H., 1951. An analysis of the operation of set in problem-solving behavior, *J. Gen. Psychol., 45*:219–244.
Guetzkow, H., and Brozek, J., 1946. Intellectual functions with restricted intakes of B-complex vitamins, *Am. J. Psychol., 59*:358–381.
Hamilton, G. V., 1916. *A study of perseverance reactions in primates and rodents,* Behav. Monogr., No. 3.
Hamilton, J. A., and Krechevsky, I., 1933. Studies in the effect of shock upon behavior plasticity in the rat, *J. Comp. Psychol., 16*:237–253.
Hanfmann, E., and Kasanin, J., 1937. A method for the study of concept formation, *J. Psychol., 3*:521–540.
Hargreaves, H. L., 1927. The "faculty" of imagination, *Brit. J. Psychol.,*

Monogr. Suppl. No. 10, Cambridge, Eng., Cambridge University Press.

Harlow, H. F., 1949. The formation of learning sets, *Psychol. Rev.*, *56*:51–65.

Hartocollis, P., 1954. Effects of alcohol on verbal performance, Ph.D. dissertation, East Lansing, Mich., Michigan State College.

Hull, C. L., 1934. The concept of the habit-family hierarchy and maze learning, *Psychol. Rev.*, *41*:33–54; 134–152.

Hunt, J. McV., and Cofer, C. N., 1944. Psychological deficit, in *Personality and the behavior disorders,* ed. J. McV. Hunt, New York, Ronald Press.

Jenkins, J. J., and Russell, W. A., 1952. Associative clustering during recall, *J. Abnorm. & Social Psychol.*, *47*:818–821.

Johnson, D. M., Johnson, R. C., and Mark, A. L., 1951. A mathematical analysis of verbal fluency, *J. Gen. Psychol.*, *44*:121–128.

Johnson, D. M., and Reynolds, F., 1941. A factor analysis of verbal ability, *Psychol. Record*, *4*:183–195.

Kendler, H. H., Greenberg, A., and Richman, H., 1952. The influence of massed and distributed practice on the development of mental set, *J. Exper. Psychol.*, *43*:21–25.

Kirkpatrick, J. J., and Cureton, E. E., 1949. Vocabulary item difficulty and word frequency, *J. Appl. Psychol.*, *33*:347–351.

Kleemeier, R. W., and Dudek, F. J., 1950. A factorial investigation of flexibility, *Educ. Psychol. Meas.*, *10*:107–118.

Lantz, B., 1945. *Some dynamic aspects of success and failure,* Psychol. Monogr., 59, No. 271.

Levitt, E. E., and Zelen, S. L., 1953. The validity of the Einstellung test as a measure of rigidity, *J. Abnorm. & Social Psychol.*, *48*:573–580.

Luchins, A. S., 1942. *Mechanization in problem solving,* Psychol. Monogr., 54, No. 248.

Luchins, A. S., and Luchins, E. H., 1950. New experimental attempts at preventing mechanization in problem solving, *J. Gen. Psychol.*, *42*:279–297.

Maier, N. R. F., 1949. *Frustration,* New York, McGraw-Hill Book Co.

Maltzman, I., Fox, J., and Morrisett, L., 1953. Some effects of manifest anxiety on mental set, *J. Exper. Psychol.*, *46*:50–54.

Maltzman, I., and Morrisett, L., 1952. Different strengths of set in the solution of anagrams, *J. Exper. Psychol.*, *44*:242–246.

Maltzman, I., and Morrisett, L., 1953. The effects of single and compound classes of anagrams on set solutions, *J. Exper. Psychol.*, *45*:345–350.

Marquart, D. I., 1948. The pattern of punishment and its relation to abnormal fixation in adult human subjects, *J. Gen. Psychol.*, *39*:107–144.

Marquart, D. I., and Arnold, L. P., 1952. A study in the frustration of human adults, *J. Gen. Psychol.*, *47*:43–63.

Oliver, J. A., and Ferguson, G.A., 1951. A factorial study of tests of rigidity, *Canad. J. Psychol., 5*:49–59.

Patrick, C., 1935. *Creative thought in poets,* Arch. Psychol., No. 178.

Patrick, C., 1937. Creative thought in artists, *J. Psychol., 4*:35–73.

Patrick, J. R., 1934. Studies in rational behavior and emotional excitement, *J. Comp. Psychol., 18*:1–22; 153–195.

Rapaport, D., 1945. *Diagnostic psychological testing,* I, Chicago, Year Book Publishers.

Rees, H. J., and Israel, H. E., 1935. *An investigation of the establishment and operation of mental sets,* Psychol. Monogr., 46, No. 210.

Robinson, E. E., 1940. An experimental investigation of two factors which produce stereotyped behavior in problem situations, *J. Exper. Psychol., 27*:394–410.

Rokeach, M., 1948. Generalized mental rigidity as a factor in ethnocentrism, *J. Abnorm. & Social Psychol., 43*:259–278.

Rokeach, M., 1950. The effect of perception time upon rigidity and concreteness of thinking, *J. Exper. Psychol., 40*:206–216.

Ross, B. N., Rupel, J. W., and Grant, D. A., 1952. Effects of personal, impersonal, and physical stress upon cognitive behavior in a card sorting problem, *J. Abnorm. & Social Psychol., 47*:546–551.

Schlosberg, H., and Heineman, C., 1950. The relationship between two measures of response strength, *J. Exper. Psychol., 40*:235–247.

Schroder, H. M., and Rotter, J. B., 1952. Rigidity as learned behavior, *J. Exper. Psychol., 44*:141–150.

Solomon, M. D., 1951. The personality factor of rigidity as an element in the teaching of the scientific method, unpublished thesis, East Lansing, Mich., Michigan State College.

Studman, L. G., 1935. Studies in experimental psychiatry: V. "w" and "f" factors in relation to traits of personality, *J. Ment. Sc., 81*:107–137.

Taylor, C. W., 1947. A factorial study of fluency in writing, *Psychometrika, 12*:239–262.

Terman, L. M., 1916. *The measurement of intelligence,* Boston, Houghton Mifflin Co.

Thompson, G. G., and Kepler, M. O., 1945. A study of the production of pleasant and unpleasant items as related to adolescent development, *J. Educ. Psychol., 36*:535–542.

Thurstone, L. L., 1938. *Primary mental abilities,* Psychometric Monogr., No. 1, Chicago, University of Chicago Press.

Thurstone, L. L., and Thurstone, T. G., 1943. *The Chicago Tests of Primary Mental Abilities,* Chicago, Science Research Associates.

Tresselt, M. E., and Leeds, D. S., 1953a. The effect of concretizing the mental set experiment, *J. Gen. Psychol., 48:* 51–55.

Tresselt, M. E., and Leeds, D. S., 1953b. The Einstellung effect in immediate and delayed problem-solving, *J. Gen. Psychol., 49*:87–95.

Underwood, B. J., and Hughes, R. H., 1950. Gradients of generalized verbal responses, *Am. J. Psychol., 63*:422–430.

Vernon, P. E., 1950. *The structure of human abilities,* New York, John Wiley & Sons.

Vigotsky, L. S., 1934. Thought in schizophrenia, *Arch. Neurol. & Psychiat., 31*:1063–1077.

Wallas, G., 1926. *The art of thought,* New York, Harcourt, Brace & Co.

Weigl, E., 1941. On the psychology of the so-called processes of abstraction, *J. Abnorm. & Social Psychol., 36*:3–33.

Welch, L., Diethelm, O., and Long, L., 1946. Measurement of hyperassociative activity during elation, *J. Psychol., 21*:113–126.

Wesley, E. I., 1953. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity, *J. Abnorm. & Social Psychol., 48*:129–134.

Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt & Co.

Youtz, R. P., 1948. The relation between number of confirmations of one hypothesis and the speed of accepting a new and incompatible hypothesis, *Am. Psychol., 3*:248–249.

Chapter 8

The Production Process: Complex Patterns and Problems

> The inventor is a fellow who doesn't take his education too seriously.
>
> —KETTERING

SELDOM, perhaps never, is the solution to a genuine problem a simple elementary response. It is a pattern of response that resolves the the pattern of the problem. In order to study the production process in its most revealing form, the preceding chapter examined situations in which relatively simple responses are produced one after another. The picture is one of production by drawing responses from a supply of responses—with more or less fluency and flexibility. The solution of complex problems includes, in addition to this relatively simple production process, the production and manipulation of more complex patterns. Production is less obvious and less copious. We infer in these cases that the production process consists largely of the organization of perceptual and cognitive patterns. The pattern of the problem situation, of the data at hand, may be broken up and new patterns organized. Simple responses may be organized into a more complex response pattern. And between the problem situation and the response pattern problem-solving aids in the form of instrumental patterns may be interpolated. A complete account of the production process as it runs off during the solution of a complex problem must include the organization, reorganization, and use of all these kinds of patterns.

THE FUNCTION OF PATTERNS IN PROBLEM SOLVING

It is convenient here, as elsewhere, to separate the pattern of response from the pattern of the problem situation to which the thinker responds.

RESPONSE PATTERNS

The important thing about a response pattern is that when one is thoroughly learned it may be produced, manipulated, and thought of as a unit, as a pattern. In checkers, bridge, and warfare the individual moves are combined into larger sequences or patterns of movement, and these, when thoroughly mastered, are produced, debated, executed, and evaluated as higher-order units. The mastery of such response patterns belongs under the psychology of skill. The psychology of problem solving is concerned with those cases in which the response pattern results, not from practice alone, but from shifting away from a practiced pattern and adopting or discovering a new pattern. Experienced players bring to each game a repertoire of patterns of moves, but a new game may require the production of a few new patterns. Likewise in warfare the standard maneuvers are well known to the planners and the forces are trained in their execution, but a new engagement may require the elaboration of new patterns of movement.

Response patterns produced during problem solving are considerably simpler than perceptual or cognitive patterns, probably because of the comparative simplicity of our effector equipment. Hence the production of response patterns bulked large in the preceding chapter, while the present chapter will consider mostly perceptual and cognitive patterns. (The response pattern in judgment is necessarily simple, for reasons that will be made clear in the next chapter.)

PERCEPTUAL AND COGNITIVE PATTERNS

Chapter 3 stated that a person has a problem when his goal-directed activity is unrewarding. Expanding this definition, we may say that perception is rewarding when it furthers the goal-directed activity of the moment, whatever the motivation and the goal may be. Perception is unrewarding, is a problem, when it does not permit the activity to go ahead. A perceived pattern, therefore, is or is not problematic depending on the motivation, whether this motivation is biological, social, or what not. In addition to such extrinsic motivation, Woodworth

(*1947*) has emphasized an intrinsic motive for perception, a "will to perceive," the goal of which is a clear perception of the situation or one that the organism can adjust to unambiguously and safely. The reward is free-and-easy organism-environment relations. A foggy or blurred environment, then, is a perceptual problem *per se* for any mobile animal. Perceptual problems of this primitive intrinsic sort are relatively rare but cannot be entirely overlooked; they may add to the difficulty of other perceptual problems. In any event, when the perceived pattern is unrewarding, new perceptual patterns are produced and tried out—as long as the motivation lasts.

At the outset, when the thinker approaches a problem, the situation is not homogeneous and unorganized. Immediate perception or transfer from previous situations usually yields some sort of patterning at the beginning. This consists of fragmentary contours, relationships, associations, and the like, or, very commonly, a general outline of the whole or a large portion of it. Further development of the pattern may proceed within the general outline, elaborating and correcting it by the addition or subtraction of parts, or by more comprehensive organization. Or some small foci of organization, such as associations by similarity or contiguity, may grow by accretion. Naturally, when the initial outline is rewarding, large-scale reorganization is not likely. And when the stimulus material is too complex to be grasped as a whole, considerable patterning of the parts will occur before any large organization is possible. In general we may say that the thinker surveys the situation, specifies the perceptual solution in terms of his motivation, and attempts to perceive the situation in this way. Perceptual reorganization by itself does not solve any problems, but it may permit the thinker to go ahead and attain a solution that he could not attain at his first view of the situation.

Instrumental or Cognitive Patterns. If the problem is complex, the above outline is too simple. The thinker often makes use of tools, techniques, concepts, principles, maps, plans, search models, and other patterns that are not parts of the problem situation nor solutions to the problem, but are produced and used instrumentally, as aids to problem solving. Tools, concepts, maps, and the like do not by themselves solve any problems, but the thinker may use them to reach solutions that he could not reach without them. Education consists in large part of the mastery of patterns that will subsequently be instrumental to the solution of life's problems.

A tool is an instrumental pattern of special significance for problem solving. This pattern includes recognition of the tool itself, knowledge of its purpose, method of application, identification of the class of objects to which it may be applied, results to be expected, and so on. Statistical techniques, surgical techniques, and the like are all tools in this respect; they may all be called instrumental or means-end patterns in that each is produced and manipulated in problem solving as a unit and as a means to an end. Mastery of a screwdriver, for example, or a technique of opinion polling implies knowledge of the results to be achieved, results that will not be achieved, procedures to be followed, etc. Man is not the only tool-using animal, we know, but he is the only animal that endows his young with tools and techniques that facilitate solution of the next generation's problems.

If the problem is a spatial one, like finding one's way through strange country, the thinker usually constructs a map or spatial pattern from which he may be able to read off his course. Similarly in dealing with logical relations he may integrate what is given into a logical pattern from which he reads off what is required. These may be called cognitive patterns if we mean by that that once the thinker has organized such a pattern he knows more about the situation than he previously did. They may be called instrumental patterns in that they are not patterns of the objective situation, nor are they response patterns; they are patterns of those aspects of the objective situation that the thinker considers important, organized in furtherance of his activity of the moment.

In general it appears that the more thoroughly the pattern is learned, as a pattern, the more useful it is in problem solving. The patterning may proceed merely to the point where similarities are noticed or, as in the experiments on semantic conditioning, where S_2 elicits the same response as S_1. It may proceed to the point where various objects or words are grouped together or designated by the same name, even though the conceptual advantages of this patterning are not realized. Children learn quite readily to say the days of the week in order, but the pattern must be learned more thoroughly before it is helpful in answering the question, What day comes after Wednesday? By the same principle alphabetizing a list of words is a step beyond rote memory of the sequence of letters of the alphabet, and children can recite the letters of the alphabet in order before they can answer the question, What letter comes before M? A well-learned pattern may be

abstracted from the various settings in which it has been encountered and manipulated *qua* pattern in thought. In some cases the solution to a problem can be "read off" an instrumental pattern once it is well formed. Some thinkers learn a principle in such a way that they can state it; others can put it to use without being able to state it. The thinker who has thoroughly mastered the principle can do both.

THE SEARCH MODEL AND THE SEARCH. Chapter 6 described a particularly important instrumental pattern, the search model. We may say that the thinker prepares for productive thought by organizing the pertinent aspects of the problem situation and the wanted solution into a partially complete pattern or search model; with this incomplete pattern as a guide he then searches the environment or his memory for something to fit it.

Probably everyone has searched for something—and psychologists are searching for answers as vigorously as anyone—but the psychology of the search has been discussed only occasionally (Johnson, *1944*). For orientation let us return to the organism in its environment and consider first the behavior of a hungry dog. If the dog explores the environment and is especially responsive to sights and smells of food, we may say that he is seeking food. It is not necessary to assume that the animal constructs a search model. The biological drive sensitizes the perceptual apparatus to promising stimulus patterns.

Consider next the case of a boy who has lost his ball. He has a problem because his first response is unrewarded; that is, he does not see the ball at first glance. His motivation to recover the ball establishes a set which may be described quite specifically. He is ready for a specified response, grasping, when he perceives a specified stimulus pattern, the ball. He has a very good search model because he knows just what his ball looks like. In this case the search model is the result of previous experience with the ball.

Next let us suppose that our boy is building a doghouse and needs a board of a certain size and shape for the floor. He has not seen the board he wants. In this case the search model cannot be a result of previous experience alone; he has to organize it himself according to the requirements of the problem at hand. Getting the size, shape, and other characteristics of the solution from his analysis of the problem situation, rightly or wrongly, he builds these specifications into an incomplete instrumental pattern, the search model, then searches the environment for a piece that completes it.

Jumping to extremely sophisticated cases, we see the same use of a search model when a mathematician tackles a difficult group of equations, when an inventor struggles to develop a better gadget, and when a scientist tries to devise a new technique or a new theory. In general the thinker gets the specifications of the solution from his analysis of the problem. He usually does not deal with these specifications one at a time, but organizes them into an incomplete pattern. With this aid he searches his memory, his library, the patent office, his old records, and other likely places for something that agrees with this pattern of specifications. Ideas are produced, not to fit the problem directly but to fit the thinker's formulation of the critical specifications of the solution to the problem, i.e., to fit the search model.

If the problem is a tough one, the search model will be reorganized as the search progresses. As the thinker moves from one point of view to another, as he digs deeper into the problem, he may specify the solution in different ways and may organize these specifications into different search models. The important function of exploratory activity is that it permits survey of the problem from different points of view.

Thus it appears that the search is a special kind of production process. All production is restricted in some way, if only slightly, as when someone names words beginning with *s*. If production is highly restricted and guided by an incomplete pattern of the wanted solution, we may call the incomplete pattern a search model and the productive activity a search.

Now all of these instrumental patterns may be brought to the problem from previous experience, as part of what Chapter 6 called the static preparation. Some, especially the search models, are developed *ad hoc* during the solution of a particular problem. Intellectual progress consists in large part of the invention of new tools, techniques, concepts, and principles by a few leaders of thought and the mastery of these by many followers. Education gives its beneficiaries the solutions to some of their problems, but more often it gives them some instrumental patterns which they may or may not be able to apply to their particular problems. To understand the psychology of thought we need to know how such patterns are organized and how they are used in solving problems.

DESCRIBING PATTERNS AND THEIR USE

Response patterns are easily described because they are objectively observable. A sequence of movements can be photographed or de-

scribed in words, and verbal response patterns can be recorded mechanically or by hand. For many purposes it is sufficient to record only the critical aspects of a response pattern, perhaps a sequence of turns or of choices of cards. That is, when only two responses are possible, the critical aspects of a behavior pattern of six moves might be represented as RLRRLL. If cards are sorted into four piles, a response pattern of six moves might be described as 134142. In the jar problems mentioned in the preceding chapter, one response pattern might be coded as $A - B$ and another as $B - A - 2C$.

Perceptual and cognitive patterns are not observable. They must be described indirectly. Much of the empirical work has been done on concepts, but other patterns may be described by the same methods.

FREE RESPONSE OR REPRODUCTION METHODS. The most obvious procedure is to ask the thinker how the problem looks to him or how he sizes up the situation. He may be asked to define a tool or concept or to state a generalization. He may be asked to diagram the significant relationships, to map the terrain, or to draw a model of the solution. The chief difficulty with the reproduction method is the difficulty of the reproduction or representation itself. Some patterns are more easily represented than others. And some people have idiosyncrasies in writing and drawing that might distort some patterns but not others. Nevertheless the free-response methods can yield useful descriptions of cognitive patterns under optimum conditions, as when someone learning a stylus maze is asked to draw the maze pattern after each trial. Definitions of concepts and other descriptive material obtained under free-response conditions may be classified by the experimenter according to his interests. Most often they are classified as right or wrong. More information can sometimes be obtained, however, when this material is classified in more specific categories. Children's answers to questions about causal relations can be classified as animistic, naturalistic, etc. Drawings can be classified as correct in general outline and erroneous in detail, and so on. Plans can be placed in categories according to an inference about the thinker's formulation of the problem. In all cases a check on the reliability of the classification procedure should be obtained by having another judge carry out the classification independently (see Chapter 2).

MULTIPLE-CHOICE OR DISCRIMINATION METHODS. Instead of classifying the thinker's freely produced activity, we can force his activity into categories by instructions or by experimental arrangements. Instead of

asking for a definition of a concept, we can ask which of five alternative definitions is the best. Here again we may merely determine whether the pattern is right or wrong or we may construct the alternatives so that choice of each one indicates the use of a different instrumental pattern. Differential responses of inclusion and exclusion are particularly informative. If we wish, for example, to describe someone's understanding of the pattern of religious discrimination in modern America, we can show him a list of activities and ask him which are part of the pattern and which are not.

HOW PATTERNS ARE FORMED AND USED IN SOLVING PROBLEMS

The solution of complex problems, if the present formulation is correct, is largely a process of producing and manipulating complex patterns: patterns of the data at hand, response patterns to fit these, and, between these two, various instrumental or heuristic patterns. We cannot explain, of course, how all problems are solved. No generalization can apply universally because each problem is a problem. Some experiments, e.g., concept-formation experiments, begin with relatively discrete material, the task being to organize a pattern to suit the experimenter. The typical problem-solving experiment begins with a complexly patterned problem situation, the task being to break it up and find or construct a solution. Some general principles can be discovered and the common sources of difficulty identified if we survey a variety of problem-solving experiments, dividing the discussion according to the nature of the problem material: verbal and logical problems, numerical problems, and spatial and mechanical problems.

VERBAL AND LOGICAL PROBLEMS

Problems that consist chiefly of the organization and reorganization of patterns of verbal relations have been favorites of the psychological laboratory. Words as stimuli and as responses are discrete, clearly marked off from each other. Letters are discrete, in the same way. But there are many letters and many thousands of words which can be related in countless patterns. The supply of words that anyone brings to a verbal problem is already patterned in various overlapping ways, as noted in the preceding discussion of the structure of the supply (see page 128). Hence verbal problem solving involves the use of previously

learned patterns of relations, the breaking up and reorganization of these, and the formation of new patterns.

THE ANAGRAM PROBLEM. The production of words from the letters ONDTERH follows the same curve as the production of words from the total supply of words (see page 192). Evidently there is an upper limit to the number of words produced under the conditions of any experiment, and the effect of this limit gives the curve its characteristic shape.

A careful investigation of the processes used in solving anagrams and the causes of difficulty was carried out by Sargent (*1940*). He used 40 groups of jumbled letters, one word to be made from each, using all letters.

Two easy problems: EJEWYRL PEEOLP
Two medium problems: UDLQII GAESAV
Two difficult problems: ETLHHA ETROS

From observations of the behavior and from the comments of 15 college students he concluded that they usually began by observing the letters as a whole, with no special plan of attack. When the problems were easy, this procedure led to an immediate reorganization of the pattern. Otherwise the subjects instituted a trial-and-error procedure, beginning with various combinations of a few letters at a time, and adding to them. Occasionally the trial and error was more systematic. Usually this trial-and-error procedure led to a sudden reorganization of the pattern, though a few subjects reported "gradual insight."

Summarizing the causes of difficulty, Sargent says:

> It would seem that the following are the features of disarranged words which make for ease of solution: presence of correct initials, final letters and combinations, cues from consonant and sound combinations, and possibly smallness of degree of displacement of letters from their proper positions. The features which make for difficulty, on the other hand, are: presence of meaningful syllables or consonant combinations not to be found in the solution word, presence of unusual letters, influence of syllabic or letter sounds different from those in the solution word, and possibly largeness of the degree of displacement of letters from their proper positions.

Sargent trained a few subjects in the talking-out method (see page 28), by which the protocol of Table 8.1 was obtained. The letter group was a rather difficult one. From this protocol and from the illustrations

TABLE 8.1. Sample Protocol Obtained by the Talking-Out Method [a]

Time in Seconds	Subject's Comments Talking-Out Method	Experimenter's Interpretation
	S-C-L-I-A-O	Reads letters. Whole method
8	C- (pause)	Begins part method
30	C-L-I, L-I, L-I-A-S, L-I, I must break the S-C-L combination	Trial and error Interference
	S-O-I, soil, S-O-I-L, Soilac, L-O, L-O-S, L-O-S-I-A-C, C-L-O-S-I-A,	Trial and error continued
75	No. S-O-A, so, soail; No. C, S-C-I-L, sclee;	
115	S-O-C-I-A-L, social!	Tried S-O combination and got solution in a flash as spelled out from S-O. Sudden reorganization
	Retrace It came in a flash from the S-O combination. There was a little interference: the S-C-L combination was hard to break up	

[a] The task was to make a word out of SCLIAO. From Sargent (*1940*).

given in Chapter 4, one would expect that the principle of embedding would operate in the anagram problem, making well-patterned letter groups more difficult to reorganize than loosely patterned letter groups. Hollingworth attacked this question directly by rearranging words, e.g., *intellectual,* into gestalted letter groups, e.g., *all in lettuce,* and into ungestalted letter groups, e.g., *alinletluetc.* But college students made as many words out of the gestalted letters as out of the ungestalted letters (Hollingworth, *1938*). However, when Devnich (*1937*) repeated the experiment with college students, recording number of errors made rather than number of words constructed, she found completely significant differences. She concluded that "configured forms convey meaning and resist being changed into other words with greater tenacity than do meaningless arrangements." But Sargent (*1940*) and Nissenson and Sargent (*1941*) again could find no differences.

The weight of the evidence at present suggests that there is no dif-

ference between well-patterned and loosely patterned letter groups in respect to ease of reorganization, but Devnich's experiment cannot be disregarded. A preliminary experiment mentioned by Nissenson and Sargent, in which the subjects were tested individually and time scores were recorded, did disclose a difference. Hence this question cannot be considered settled; there must be some unspecified variations in these experiments, perhaps in the kinds of scores recorded, that produce such inconsistent results.

A series of anagram problems can be arranged so that the first few establish a set that carries over to the solution of later problems (Rees and Israel, *1935*). Such a task-induced set can make the later problems more difficult than they would otherwise be (Luchins, *1942*).

The order in which words are read also influences the way they are organized in patterns, as in an experiment briefly described by Cofer (*1951*). He gave his subjects lists of words, like the following, with instructions to find the one that does not belong with the others.

SKYSCRAPER TEMPLE CATHEDRAL PRAYER

Because of our left-right reading habit, the reader is likely to group the first three words as buildings, hence prayer is the one that does not belong. Another order could be:

PRAYER TEMPLE SKYSCRAPER CATHEDRAL

The first two and the last could be grouped as religious words, hence it is skyscraper that does not belong. By presenting the four words in all possible sequences, Cofer showed that words seen first have more weight in determining the pattern than words seen last.

THE CONCEPT-FORMATION PROBLEM. The formation of a concept does not directly solve any real-life problem, but concepts are widely used in the communication of information and principles that are instrumental in solving real-life problems. Furthermore people often develop private concepts that they use in their own thinking even though they may not communicate these to anyone else. Thus the formation and use of concepts is an important topic for education and psychology, both because the result, mastery of the concept, is of practical value and because the process of concept formation is of considerable theoretical significance. In addition an experiment on acquiring a concept often turns out to be an experiment on solving a problem, thus illuminating a special case of problem solving.

The term concept has had many meanings in its long and troubled history, but present interest is centered on the general or class concept, an abstract or cognitive pattern of the common characteristics of a number of different objects, events, or ideas. A concept may be described as a spherical pattern, with a center of precise meaning and a periphery of connotations. Socially accepted concepts are always designated by a communicable symbol; private concepts may or may not be so designated.

Like other patterns, concepts can be studied by a free-response method or by some kind of objective discrimination method. A child's concept of dog can be tested by asking him to describe a dog. His description can be scored "accurate," "too general," "abstract," "concrete," "irrelevant," etc., with a fair degree of interscorer agreement. In fact items of this kind are used on intelligence tests and scored with good reliability. When the description avoids irrelevancies and includes only the socially accepted meaning, we call it a definition. The description may also be pictorial, as when a child is asked to draw a triangle or a college student is asked to draw a neuron. The descriptions are influenced, of course, by the subject's skill in verbal and pictorial techniques as well as by his mastery of the concept.

A free-response method that is not very free because it is limited by the possibilities of the situation presents the subject with a variety of objects with instructions to "put those together that belong together" (see Chapter 7). The subject's conceptual system may be inferred from his grouping, and in addition he is usually asked for a verbal explanation of his grouping.

The discrimination method has several variations. The subject may, for example, be shown a variety of objects, either serially or simultaneously, and asked which are cylinders. Many experiments on the acquisition of concepts have proceeded by this method, using rewarded practice to attach a new name, e.g., *dax,* to a new class of objects or drawings. Progress is indicated by the number of objects correctly named. Names are avoided and the problem becomes suitable for nonverbal animals if a nonverbal response, such as turning or approaching, is attached to positive instances and not to negative instances. A variant of this method is the oddity method, in which several objects are presented, all but one of which are similar in some respect, and the odd object is to be identified by an approach response,

by underlining, or by some other obvious indicator. The well-known analogy problem is similar.

Any of these methods can be used to test whether the subject is using the concept standardized by the culture or the experimenter or to determine the personal concept the subject is using. The free-response method, however, is somewhat more appropriate for the later assignment. In general, experimental psychology has been interested in the mastery of standard concepts, scored right or wrong by the experimenter, while child psychology and abnormal psychology have been interested in describing whatever private concepts the subject may actually hold (see Vinacke, *1951*). Educational psychology has been interested in both problems.

If the sample tested contains subjects who vary widely in mastery of a concept, both free-response and discrimination methods will yield similar results. But the two methods do not test exactly the same achievements, and some can pass one test but not the other. This disagreement between the different criteria of mastery showed up in a pioneer experiment by Ach (*1921*). Hull (*1920*) and Smoke (*1932*) found, also, that laboratory subjects can often discriminate one class of designs from the others without being able to describe or draw the common characteristic of the class. An opinion poll (American Institute of Public Opinion, *1939*) showed that ordinary citizens can classify public figures as radical, liberal, and conservative with considerable accuracy, yet may be quite unable to give adequate definitions of these concepts. We should not, however, jump from these results to the conclusion that a definition is the best test of the mastery of a concept. Many people have memorized definitions that they cannot use for classifying objects. As in any other area of science, the results are determined in part by the method, and the best results come from a convergence of two or more methods.

The first question about concept formation is the question of motivation, and the answers at present are only guesses. In the daily routine of social life, concepts are learned for purposes of communication, rewarded by progress toward many diverse goals. In school, concepts are deliberately studied for purposes of prestige, social approval, and many others. It is worthwhile also to consider the outrageous hypotheses that these extrinsic motives are complicated by an intrinsic motive for conceptualizing *per se*. Just as man, with a powerful perceptual apparatus, may be endowed with a "will to perceive," so may he, since he pos-

sesses an extraordinary capacity for abstraction, be endowed with a will to conceive. If there is a natural tendency to clear up a confused perceptual environment, so there may be a natural tendency to impress order on a confused intellectual environment. The reward is in one case a stable environment that one can readily and safely get around in, and in the other case a stable system of concepts that enables one to find his way in the intellectual world. If there were no natural tendency to divide up the world into concepts, the human organism would soon invent one, because the labor-saving value of a system of concepts is enormous and any moderately useful system will be continuously reinforced.

The second question about concept formation, how it proceeds, receives several answers. Concepts can be acquired by a simple type of learning, by complex learning or problem solving, or by some combination of these. Serial presentation of unfamiliar stimulus objects with instructions to learn their names, as in the classic experiments of Hull (*1920*) and Kuo (*1923*), favors a simple gradual kind of acquisition. The principles of stimulus generalization and differential reinforcement may be sufficient to account for these simple cases, especially if the stimulus material varies in only one dimension (Spence, *1937;* Buss, *1950*). Concepts of a higher level of abstraction can be accounted for theoretically by a configural extension of these principles (Welch, *1947a*). Serial presentation with instructions that emphasize discovery of the common characteristic favors a more active search or problem-solving process (Smoke, *1932;* Drever, *1934;* Reed, *1946a*). Simultaneous presentation of the stimulus objects with instructions to find the common characteristic or to "put those together that belong together" reduces the role of practice and memory and makes the task a purer problem in discovery (Heidbreder, *1949;* Reed, *1950*).

During active search for the common characteristic tentative solutions or hypotheses are produced by transfer from other situations, by observation of similarities and differences in the present situation, and by summation of subliminal response tendencies. It is possible under some conditions with suitable instructions to record the concepts the subject thinks of, or at least reports, and the rate of such evocation of concepts is closely related ($r_{tet} = .85$) to successful solution (Marks and Ramond, *1951*).

Promising concepts or hypotheses thus produced are verified or rejected as experience accumulates. Information contrary to a hypothesis

(negative instances) is not used as effectively as one might expect from a purely logical analysis (Smoke, *1932;* Hovland and Weiss, *1953*). Often a concept is not rejected entirely but is reorganized, some features being added and others subtracted. Unfamiliar material is often assimilated or even forced into familiar categories (English, *1922*), sometimes with a qualification, as when a child describes an elephant as a horse with two tails. Organization and reorganization of concepts is a lifelong activity, but assimilation of new data to the old categories seems to be more comfortable than reorganization of one's concepts to fit the new data.

Many concepts and their names are acquired from the context in which they occur, either incidentally, as one reads rapidly along, or deliberately, as when one stops to puzzle over the unfamiliar word. Gibbons (*1940*) has shown how difficult this process is by asking college students to read sentences that contain an unfamiliar word, and to try to define this word:

In the beginning the teacher traveled from one locality to another to meet the students, thereby bringing into existence the *itinerant* school master.

Assuming that teacher and school master are synonymous, the cues to the meaning of *itinerant* are *traveling* and *thereby*. Out of 234 college freshmen who undertook this task, 156 or 67 percent were able to give an acceptable meaning. Of the other 78 who failed, over half had no idea at all of the meaning and the rest gave such replies as a research school master, an immoral school master, a rural school teacher, a country teacher.

A similar method for studying the same problem has been described by Werner and Kaplan (*1950*). They use artificial words and give several contexts for each. One problem was to discover the meaning of *corplum* from these sentences.

A corplum may be used for support.
Corplums may be used to close off an open place.
A corplum may be long or short, thick or thin, strong or weak.
A wet corplum does not burn.
You can make a corplum smooth with sandpaper.
The painter used a corplum to mix his paints.

The sentences are shown to the subject one at a time, and the subject is asked what *corplum* means and how he discovered the meaning. Older children can do this task better than younger children, but, more

than that, the experimenters, by analyzing the children's responses, have classified the processes by which the meaning is discovered.

CONCEPTUAL SYSTEMS AND THEIR USES. Most of the laboratory experiments on concept formation have dealt with rather simple concepts, but the acquisition of high-level concepts and their hierarchical arrangement are equally important for the psychology of thought. The most thorough investigations of the acquisition of concepts at various levels of abstraction and their use in subsequent problems have been carried out by Welch (*1940*) and Welch and Long (*1940, 1940a*). The second of these (Welch and Long, *1940*), using 103 children 42 to 83 months in age, is most convenient for illustration of method and results.

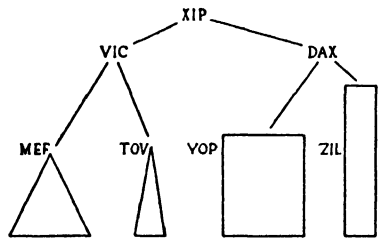


FIGURE 8.1. Sample of materials used by Welch and Long (*1940*) for study of higher levels of concept formation in children.

To investigate the learning of names at the object level, i.e., no abstraction, the experimenters taught the children by the prompting method that a certain square was called *Vim,* a certain hexagonal bead *Jid,* and a round bead *Quap.* These names were readily learned by most children, evidently by a simple sort of associative learning (see Figure 8.1).

To investigate the learning of names for the first level of abstraction two triangles were used. One with a broad base was called *Mef,* and one with a narrow base was called *Tov.* The class of triangles was called *Vic.* Two rectangles also were used, one with a broad base called *Yop* and one with a narrow base called *Zil.* The class of rectangles was called *Dax.* Learning these class names, *Vic* and *Dax,* at the first level of abstraction, went fairly well, 66 percent reaching the criterion for triangles and 67 percent for rectangles.

To investigate the learning of names for the second level of abstraction, those who passed the first-level test were taught that all triangles and rectangles, that is, all objects called either *Vic* or *Dax,* were to be called *Xip*. Only 29 percent were able to learn this more difficult concept to the experimenters' criterion.

The experimenters conclude that "it is easier for a child to learn to identify three objects than to learn to identify a hierarchical structure of two objects and their class." Furthermore, "mastering a second hierarchical structure of four objects and three class concepts proved to be a still more difficult task."

An important part of this study which came later was the test of the children's ability to use these names in following commands. They were given such commands as: Put all of the *Quap* on this side of the table and all of the *Jid* on the other side (object level). Put all the *Vic* together and then put all the *Dax* together (first level of abstraction). Put *Xip* in the middle of the table (second level of abstraction). Obeying commands at the object level was easy, of course, but many errors were made in obeying commands at the first level of abstraction. Obeying commands at the second level of abstraction turned out to be easier than expected, probably because of the practice on the first level.

Discovery of a Principle. A principle is a concept in the sense that events can be classified as instances of a principle or as not instances of this principle. But a principle is more than a concept; it states a relationship between concepts and thus can be used to arrive at new conclusions or deductions or to guide new activities. Therefore, in addition to the methods of description by free response and by discriminating situations to which a principle applies from situations to which it does not apply, we have a new method: application of the principle to new situations to obtain new knowledge. An experiment that uses the third method is *ipso facto* a problem-solving experiment.

If the subject's incidental behavior is dependent upon his acquaintance with a principle, it may be feasible to test for the latter by observing the former. Dixon (*1949*) observed nursery-school children when faced with an illusory contradiction of their expectation of size constancy (what had looked big now looked small), taking surprise, confusion, spontaneous verbalization, etc., as indicators of familiarity with the apparently contradicted relationship. He reported good agreement between two observers and between two presentations, hence the the changes in incidental behavior must have occurred; and it is

likely, though not certain, that the contradiction of the expectation produced the changes in behavior. This is an additional possibility, therefore, for studying children's implicit or unverbalized generalizations.

Some principles are learned casually, perhaps thoughtlessly, as children learn, while bouncing a ball, that the angle of incidence equals the angle of reflection. Some are learned from other people, either in casual conversation or by serious study from books and teachers in school. And some are discovered by serious problem solving.

Laboratory experiments in which the problem is the discovery of a principle indicate, as in other kinds of problem solving, an alternation between the production of hypotheses and the testing or evaluation of these. Not much is known about the actual production of these hypotheses or tentative principles, but probably the production process resembles the production of concepts. Surely the thinker performs some sort of analysis of what is given and what is required. This analysis may lead to the organization of a search model which guides production of hypotheses. Heidbreder's (*1924*) study emphasized a difference between participant and spectator behavior. The former was an overtly active testing of hypotheses; the latter was a period of no obvious activity—presumably thoughtful waiting—often followed by the production of fruitful hypotheses. It may be that the testing of hypotheses and the production of hypotheses cannot be done simultaneously. The former, the more observable phase, may have to wait for the latter, which is less observable. We shall return to spectator behavior later in this chapter, under the heading of incubation.

Other Instrumental Patterns. Concepts and principles do not, by themselves, solve any problems. They are important because they can be used in the solution of many different problems. Tools and techniques, likewise, are valuable only because they can be used in solving problems. Some tools and techniques, however, are useful in the solution of only one particular problem. Mastery of tools and techniques could be tested as mastery of principles is tested: by asking that they be described in words or by drawing, by asking for discrimination of problem situations to which they apply from those to which they do not apply, or by asking for their direct application.

The search model is an instrumental pattern that has not been studied in this way because a search model is like a tool that is specially designed for only one task, a search for only one solution. But when

someone is working on a complex problem, we could study his search model by asking him to describe what kind of answer he is looking for, by asking him which of a list of solutions would fit his search model and which would not, and by noting what solutions he actually produces. (Note also the proposed experiment on preparation for perceiving on page 189.) In general, we need evidence on the construction of maps, plans, programs, and other patterns that do not directly solve any problem, but are instrumental to the solution of genuine problems.

Factors Affecting the Rate of Acquisition of Concepts and Principles. The evidence on rate of acquisition comes from experiments in which the independent variable is amount of practice, as measured by number of trials or instances or promptings. Modern communication theory suggests a more sophisticated independent variable, the amount of information conveyed by the trials (Hovland, *1952*).

One clearly identified difficulty in organizing concepts and other abstractions is their very abstractness. The Welch and Long experiments showed that concepts of directly perceived drawings are easier to master than concepts of a higher level of abstractness. This principle can also be demonstrated by a slightly different procedure. When college students are shown drawings of various objects and asked to memorize their class names, concepts of concrete objects are mastered most readily, with concepts of spatial forms next, and numerical concepts last (Heidbreder, *1946, 1947;* Heidbreder, Bensley, and Ivy, *1948;* Wenzel and Flurry, *1948;* Grant, *1951;* Grant and Curran, *1952*). Hines (*1950*), using Reed's procedure (see Table 8.2), found that abstract concepts, like discipline and religion, required about twice as much practice as concrete concepts, like tree. According to Heidbreder, perception of concrete objects is the dominant mode of cognitive reaction, related to the phylogenetic priority of the locomotive and manipulative capacities of the organism. Conceptualization is therefore easier the more it resembles the perception of concrete objects.

If concepts of concrete objects are most readily mastered, they will be most readily available for use in problem solving. Hence if several alternative conceptualizations are possible, preference will be given to concrete concepts. Many instances from the history of thought and the diffusion of knowledge are consistent with this principle. Concepts like germs, vitamins, atoms, and molecules catch on quickly, while explanations in terms of molecular structure are hard to grasp, and systems of

mathematical relationships never do become popular with the general public. Teachers commonly make numerical concepts easier by translating them into spatial or graphic form, and still easier, if possible, by constructing visible models. Later studies by Heidbreder and her collaborators indicate that this principle is not an extremely prominent one, but is often obscured by other tendencies. (Heidbreder, *1948;* Heidbreder and Overstreet, *1948*).

The abstractness that makes abstract concepts hard to acquire has the same effect on acquisition of principles, except that the effect is not so prominent—perhaps because all principles are rather abstract at the outset. Long and Welch (*1941, 1942*) had children master certain principles of arrangement of stimulus materials in order to make a light turn on. Then the children had to use these principles with other kinds of material. Horizontal generalization was quite successful; that is, a principle mastered on one kind of material was readily applied to other kinds of material. But the principles were applied more readily to three-dimensional objects than to pictures, words, or letters. And principles of higher levels of abstractness were in general more difficult than principles of the object level.

Other difficulties affecting the formation of concepts are in the stimulus situation and the context of the material from which the concept is to be organized. It is obvious that any arrangement of the situation which helps to isolate the common characteristic from irrelevant characteristics facilitates conceptualization. Hull (*1920*) demonstrated long ago that when the common characteristic of meaningless drawings is emphasized, e.g., outlined in red, it is more easily identified, and skillful teachers have always made use of this trick. In reverse, one would expect that embedding the common characteristic in an inclusive pattern would make it harder to isolate. Gengerelli's (*1927*) study of mutual interference showed that when perceptual material is organized one way, shifting to another way of organizing it becomes unusually difficult. Davidon (*1952*) found that subjects allowed to manipulate the materials of a concept-formation experiment do better than those who have to leave the materials as they are. Grouping the materials presumably fixates those concepts already discovered and prevents subsequent confusion. An extension of these findings to the mastery of principles would assert that principles are easier to discover when those variables that change together in a complex sequence of events are pointed out or illuminated or separated in

some way from irrelevant variables. This rule also has long been used by skillful teachers.

A series of experiments by Reed (*1946a, b, c*) illuminates the effects of several other factors on rate of acquisition of concepts. By varying instructions to the subjects, Reed showed that a set to discover concepts was more effective than a general set to learn correct responses. (See also Ewert and Lambert, *1932*.) Concepts based on superficial similarities, usually inconsistent or nonlogical, were acquired more slowly than logically consistent concepts. When the correct concept was buried in irrelevant material, mastery was made difficult, especially when the irrelevant material could be confused with the correct concept (see Table 8.2).

TABLE 8.2. Selections from Materials of an Experiment in Learning Concepts [a]

	Group of Stimulus Words (on card shown to subject)				Name (to be learned)
1.	horn	leaf	monkey	debt	kun
2.	brook	leave	claim	precious	vor
3.	roses	suit	juice	plum	yem
8.	fame	ought	tiger	saucer	kun
11.	lover	borrow	flower	point	vor
18.	line	people	elephant	sound	kun

[a] From Reed (*1946a*).

Reed also compared the simultaneous form of presentation with the more usual serial form. The number of correct concepts discovered was greater for the former, presumably because when the thinker can see both the supporting and nonsupporting evidence for any hypothesis in front of him, he can draw a more accurate conclusion. The subjects spent considerable time studying the material displayed simultaneously, hence this form of presentation is not necessarily more economical. When the serial form of presentation is used, distributed practice seems to be more economical than massed practice (Oseas and Underwood, *1952*). At least the intertrial intervals of 15 to 60 seconds were more economical than the 6-second interval.

THE PATTERN OF PERSONALITY. Some interesting information on organizing trait names into patterned descriptions of personality comes from experiments by Asch (*1946*). He used a free-response method,

giving college students a few adjectives descriptive of people, from which they were to write paragraphs giving their impressions of the people to whom the adjectives refer. By way of illustration: when these terms were read to the subjects, energetic, assured, talkative, cold, ironical, inquisitive, and persuasive, one subject wrote the following paragraph:

> He seems to be the kind of person who would make a great impression upon others at a first meeting. However as time went by, his acquaintances would easily come to see through the mask. Underneath would be revealed his arrogance and selfishness.

Another wrote this paragraph:

> Possibly he does not have any deep feeling. He would tend to be an opportunist. Likely to succeed in things he intends to do. He has perhaps married a wife who would help him in his purpose. He tends to be skeptical.

Evidently, though the subject hears a sequence of discrete trait names, his impression is not a sum of discrete impressions. He organizes the data into a more or less consistent pattern. It is the psychologist's job to describe how this organizing process occurs. Beginning this job, Asch used two lists of words, both of which contain the words calm and strong:

A. kind, wise, honest, calm, strong
B. cruel, shrewd, unscrupulous, calm, strong

List A was read to 38 subjects and List B to 41 subjects, all of whom were asked to write synonyms for each term. The instructions were, "Suppose you had to describe this person in the same manner, but without using the terms you heard, what other terms would you use?" Some of the most revealing responses are collected in Table 8.3.

This table shows quite clearly that "calm" has different meanings within the context of List A than within the context of List B. Synonyms for "strong" show the same variation. Within the context of List A some of the synonyms for "strong" were fearless, helpful, just, forceful, courageous, and reliable. Within the context of List B some of the synonyms for "strong" were ruthless, overbearing, overpowering, hard, inflexible, unbending, and dominant.

TABLE 8.3. Frequency of Use of Various Words as Synonyms of "Calm" When "Calm" Appears in Different Lists of Words [a]

Synonyms	List A kind . . . strong	List B cruel . . . strong
serene	18	3
cold, frigid, icy, cool, calculating, shrewd, nervy, scheming, conscienceless	0	20
soothing, peaceful, gentle, tolerant, good-natured, mild-mannered	11	0
poised, reserved, restful, unexcitable, unshakable	18	7
deliberate, silent, unperturbed, masterful, impassive, collected, confident, relaxed, emotionless, steady, composed	11	26

[a] From Asch (*1946*).

Several other conclusions could be drawn from these simple but instructive experiments. Two are of especial significance for the organization of patterns. By varying the order of the words in the lists it was discovered that the first word is particularly important in organizing the list. And Asch noted that his subjects supplemented the data, importing new adjectives and integrating them with the given adjectives.

PROBLEMS OF LOGICAL RELATIONS. Forming concepts, principles, and patterns of personality traits may be called inductive tasks in that the particulars are organized into patterns. But we do not observe these cognitive patterns directly; we test for the pattern by asking the thinker to use it in some particular way, that is, by a deductive process. And, in reverse, problems that are called deductive, e.g., the syllogism, always include an inductive step because the particulars of any problem have to be organized before any conclusions can be produced. It should be possible to invent an objective method for classifying logical problems as predominantly inductive or predominantly deductive in respect to the principal source of difficulty. The application of

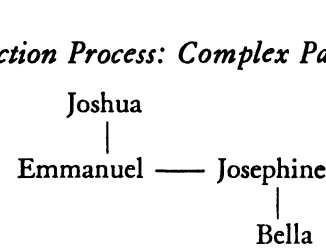
factor analysis to this purpose has not been completely successful (See page 398). At present the distinction between inductive and deductive problems is made by subjective or logical examination, and really cannot be taken very seriously. Some psychologists call all such problems reasoning problems.

Some logical problems—and these might plausibly be called deductive—require the discovery of logical relations or the rearrangement of logical patterns in such a way that a conclusion can be attained or a question answered. Consider these four problems:

1. John is taller than Mary, and Enoch is shorter than Mary. Which boy is taller?
2. Emmanuel is shorter than Joshua, and Josephine is taller than Bella. If Josephine and Emmanuel are the same height, is Bella the same height as Joshua?
3. Louise is taller than Richard, who is the same height as his brother Enoch. Rachel is shorter than Richard but taller than her twin brother Florio. Which girl is taller?
4. If all chigs are falten,
   And all gorils are falten,
   Can we be sure that some gorils are chigs?
5. No *x*'s are *y*'s;
   No *y*'s are *z*'s;
   Therefore no *z*'s are *x*'s. *True or false?*

The first problem is so easy that people often get the answer immediately; perhaps it should not be called a problem. The other problems are more complicated. People often prepare for these complicated problems by rereading, familiarizing themselves with the terms before undertaking any rearrangement. We know that familiarity is of some importance because Wilkins (*1928*) found that syllogisms written in familiar terms are somewhat easier for college students than syllogisms of the same logical form written in unfamiliar terms, as in No. 4, or in terms of letters, as in No. 5.

The production process in working these problems consists of the production of meanings of the terms or concepts, associations with these, relationships between terms, rearrangements of the data, etc. Students of logic may, of course, recall similar problems from previous experience and reach a conclusion without much thought. Many thinkers report that they do not seek a conclusion directly, but try first to arrange the data into a pattern, like the one here—which we



may call an instrumental pattern—from which they can read off the desired conclusion. Euler's circles, often used in the solution of syllogisms, may be considered instrumental patterns in this sense. Some people construct such patterns in imagination, or on their fingers, and some actually draw them on paper.

The data are not unordered to begin with. And some dispositions of the data are easier to rearrange than others. Knowing that John is taller than Mary, for example, one has to reverse the relation between Enoch and Mary so that it reads "Mary is taller than Enoch" before the two relations can be put into one pattern. Such reversal, or conversion, does take a little effort, and a problem requiring a conversion is more difficult than one in which the relationships are given in a more straightforward form (Burt, *1919*).

Both No. 2 and No. 3 are harder than No. 1. No. 2 is harder because the instrumental pattern, from which the solution is read off, is more complex, includes additional relations. No. 3 is harder than No. 1, not because the instrumental pattern is more complex—it is exactly the same—but because the problem situation is more complex, containing irrelevant data from which the relevant data must be disentangled.

Disentangling what is logically relevant to the conclusion from the irrelevant data is not so easy as it might appear. In the two premises of No. 5, for example, there are no data relevant to the conclusion. But each premise has a "no" in it, and so does the conclusion. A surprisingly large number of people will accept the negative conclusion because of the negative impression created by the premises. This "atmosphere effect," as it is called, has been clearly demonstrated in experiments on college students using syllogisms of a variety of types (Woodworth and Sells, *1935;* Sells, *1936;* Sells and Koob, *1937*).

NUMERICAL PROBLEMS

Number Names and Combinations. The simpler number concepts, 1, 2, and 3, are probably acquired by a simple kind of learning,

like the nonsense concepts studied by Hull (*1920*)—though by definition some degree of abstraction is involved. Number combinations, such as $2 + 3 = 5$, are also acquired in a simple repetitive way. And, as in the case of verbal concepts, children may learn the names of numbers and be able to count without being able to discriminate four objects from five. They may learn to say $2 + 3 = 5$ without being able to apply this relationship to pertinent situations.

Modern number systems are highly generalized and consistent. The relationships between small numbers within this system are therefore few and relatively easy to learn. Even at the second-grade level children can study only about half the addition facts and generalize the remainder. When well taught this generalization method is superior to drill methods (Olander, *1931;* Overman, *1931;* Thiele, *1938;* Swenson, *1949*). Thiele, for example, compared the progress of several hundred second-grade children learning addition combinations by drill methods with experimental groups who were given opportunities to discover for themselves the general relations between numbers and to extend such generalizations to larger numbers. Both groups spent the same amount of time on arithmetic. The experimental groups improved much more than the control groups and their superiority was most marked on number combinations that are usually considered difficult. The experimental groups were superior also in applying their skill to unfamiliar addition problems.

This generalization, like most others, is too general. Anderson's (*1949*) large-scale experiment with fourth-grade children shows up some of the details. Teaching procedures that emphasized generalization and discovery of relationships were best for children of high ability and initial low achievement, but drill procedures were just as good and perhaps better for children of low ability and high initial achievement. This illustrates in a new way the old truism that to determine the most effective teaching method we must know the principal causes of difficulty. We shall shortly list the principal causes of difficulty in arithmetic problems.

The result of this conceptualization process, if all goes well, is the development of a highly abstract scale of quantity, with points along it clearly denoted by integers, and points between these denoted, usually with less clarity, by fractions. The scale is not perfectly uniform, to be sure. To most people 5 and 10 are somewhat different from 4 and 11 (see Chapter 10), 0 has a special status, and the very large

numbers, approaching infinity, are rather vague and seldom used. But even though the number scale is imperfectly generalized, many relationships between numbers are well known to most adults. Hence the number system becomes a complex pattern of relationships, an instrumental pattern that can be applied to many objects and aspects of the environment and used for the solution of many problems.

Some numerical concepts are not so precise as 7 and 43. Semiquantitative or indefinite concepts, such as "a few" and "many," are in common use. Research on school children by Brotherton, Read, and Pratt (*1948*) shows considerable variability in the use of six indefinite number concepts to represent numbers from 3 to 33. By and large, however, terms like "hardly any" and "few" are used for small numbers of things; terms like "many" and "a lot" are used for large numbers. The association between the indefinite and definite number concepts in Grades 6 to 11 may be represented approximately by a contingency coefficient of .73 (Johnson, *1950*).

Methods of Solving Arithmetic Problems. It is possible to observe people working arithmetic problems, asking questions now and then, and get some information about the processes by which they reach their answers. This method has been used by Brownell and his students (*1941*) with first- and second-grade children. Each child was interviewed by his teacher. A simple addition or subtraction problem was given to the child, and he was asked to "think out loud." The activities of the children were classified by the teachers according to the following code:

No attempt; says, "I don't know," or remains silent without offering an answer.

Guesses; "takes a chance"; is satisfied with incorrect answer, or attempts to correct by guessing again.

Counts sums or remainders by starting with 1 or ending with remainder.

Partly counts; in addition, does not begin with 1, but counts to one of the numbers given; in subtraction, the procedure is the same as counting.

Solves in one or another of several ways:

1. Reverses order of numbers (thus, changing a given $1 + 3$ to a better known $3 + 1$)
2. Relates the addition and subtraction facts; gets the answer for an addition fact by thinking of the corresponding subtraction fact (thus, gets $6 - 2 = 4$ from $4 + 2 = 6$)

3. Adds or subtracts from doubles (thus, $3 + 2 = 5$, because $2 + 2 = 4$, $+ 1 = 5$; or $3 + 3 = 6$, $- 1 = 5$)

Uses some other more familiar fact in a way not described above.

Meaningful recall; the child "knows" the answer, and states it at once, confidently, without use of roundabout procedures.

Teachers seemed to be able to use these categories, though it would be well, of course, to have some measure of agreement between observers.

This type of analysis was used with 40 first-grade children and 60 second-grade children. Each was given four addition problems, such as $4 + 6$, and four subtraction problems, such as $11 - 3$. The teachers' classification of their children's activities in response to these problems is shown in Table 8.4.

TABLE 8.4. Percentage Frequencies of Procedures Employed by First- and Second-Grade Children Working Simple Arithmetic Problems[a]

Procedure	Percentage in Grade I	Percentage in Grade II
No attempt	5	3
Guesses	27	9
Counts	—	—
Partly counts	6	7
Solves	8	21
Meaningful recall	54	60

[a] Data from Brownell, *et al.* (*1941*).

From experiments of this kind it is possible to construct a description of the mastery of arithmetic in the lower grades which turns out to be very similar to the usual description of the mastery of nonsense syllables. The material is initially meaningless, unorganized. Combinations are often learned in a mechanical, meaningless way, and as such are soon forgotten. But children also struggle to put some sense into the arithmetic they have to learn, to make use of whatever relationships they can discover. Some methods of teaching facilitate this organizing process more than others, as noted above. Well-organized material is remembered better and is more available for later use than disparate combinations learned mechanically. Later, after the simpler arithmetic combinations are organized in some way, they become habitual or mechanical, and now can be used with high speed and

accuracy. A later study by Brownell (*1944*), using the approach described above, showed that third-grade children, working simple multiplication problems, do less than half of them mechanically and are forced to resort to counting, partial counting, reversal, and other problem-solving techniques for a large number of problems. But in the fifth grade most problems, about 87 percent, were solved by immediate meaningful recall; problem-solving processes were quite infrequent.

These studies have used the technique of subjective analysis described in Chapter 2. There is much to be learned by this technique of investigation, especially since a variety of arithmetic problems can be easily constructed to fit the specifications of any hypothesis, and the responses can be easily recorded and compared with objective data. Ways of measuring the reliability of this kind of analysis are described in Chapter 2.

Sources of Difficulty. A more objective technique of investigation, and a much more common one, is a simple comparison of different kinds of problem in respect to number solved. This technique yields evidence on the relative difficulty of different types of problem and permits inferences about the causes of the difficulty and possibly about the psychological processes involved.

1. Certainly one factor that makes some problems harder than others is the difficulty of the simple arithmetic combinations involved in the solution. The relationship cannot be close, however, because of the possibility of generalization, or transfer, and the influence of teaching methods on amount of generalization.

2. Another factor in the difficulty of the problem is the familiarity of the setting of the problem. Brownell and Stretch (*1931*) investigated this factor by constructing four tests, phrasing the same problems in four different ways. Form A was written in terms familiar to children, e.g., children making posters. Form B used less familiar terms, e.g., soldiers grooming cavalry horses. Form C used terms even less familiar, and Form D used artificial terms, e.g., *shulahs* destroyed by *planti*. The experimenter's judgments of familiarity were confirmed by teachers' ratings. When problems made up in these four forms, carefully equated otherwise, were given to 256 children in several fifth grades, the percents correct were (A) 64, (B) 58, (C) 57, (D) 51. These differences are not large, considering the large differences in familiarity, but they are significant. They indicate the same small

effect of familiarity as that found by Wilkins in her study of college students solving syllogisms (see page 244). Further analysis of these four types of arithmetic problems showed that the principal reason for these children's errors was not in computation, but more often in choosing the wrong arithmetic process, e.g., adding when they should multiply. The effects of unfamiliarity are stronger in children of low arithmetic ability.

3. A third factor in the difficulty of arithmetic problems may be their artificial or impersonal nature. Problems of personal significance that children bring to school and problems that develop out of schoolroom activities seem to be easier than similar artificial problems. Furthermore, practice in solving such problems seems to be more effective than practice in solving the usual textbook problems. The evidence on this topic, and on arithmetic problem solving in general, has been summarized by Wilson (*1950*). The relative ease of personal problems may be due in part to familiarity of the setting and in part to more persistent motivation; but in experiments by Marks (see page 165) and Marks and Ramond (see page 234) the real-life problems were harder—presumably because the personal involvement disrupted the analytic attitude.

4. We have mentioned earlier the initial survey of the problem and the necessity in complex problems for reorganization of the problem material. Subjective reports point to the importance of this reorganization of the data, and there is objective evidence from experiments with logical problems that the reorganization is one source of difficulty.

If reorganization of the material of arithmetic problems is necessary, it should be possible to arrange problems in which this is the principal source of difficulty. Duncker (*1945*) has done just this, using the problem: Why are all six-place numbers, of the form 276,276, 591,591, 112,112, divisible by 13? It is a difficult problem for high-school and college students; few can solve it without aid. To show the importance of reorganization of the problem material, Duncker gives some groups aids that facilitate reorganization and other groups aids that do not facilitate reorganization. The aids given and the percentages of resulting solutions are given in Table 8.5.

Evidently only the first two aids were really helpful. (The numbers in these groups are small, but calculation of chi square by the present writer shows that the differences between the percentages are significant well beyond the 1 percent level.) It is always difficult, of course,

TABLE 8.5. The Effects of Various Aids on Percentage of Solutions to the "13" Problem [a]

Aid	Number of subjects	Percentage of solutions
"The numbers are divisible by 1001"	22	59
"1001 is divisible by 13"	10	50
"If a common divisor of numbers is divisible by 13, they are all divisible by 13"	13	15
"If a divisor of a number is divisible by p, then the number itself is divisible by p"	22	14
"Different numbers can have in common a divisor which is in turn divisible"	10	0
"Look for a more fundamental common character from which the divisibility by 13 becomes evident"	13	15
No aid	26	8

[a] From Duncker (*1945*).

to specify the effect on the subject of statements made by the experimenter, but the first two statements were made up according to a hypothesis about problem solution while the others were not. The facts therefore add some support to the hypothesis that solution is aided most by breaking up the numerical pattern initially seen as *abc, abc,* and reorganizing it as a 1001 pattern.

Another way to emphasize the 1001 aspect of the numerical pattern is to place two illustrations together in a suggestive relation, viz., $\begin{matrix}276{,}276\\277{,}277\end{matrix}$ as contrasted with $\begin{matrix}276{,}276\\591{,}591\end{matrix}$. Duncker did this with the expected results, but the number of cases is too small to be taken seriously.

Duncker's 1001 problem can only be solved by a reorganization of the numerical pattern. Wertheimer (*1945*) has collected many solutions of a numerical problem that could be solved either by a blind piecemeal procedure or by a structural reorganization. The problem is to get the sum of a series of numbers:

$$1 + 2 + 3 + 4 + 5 = ?$$

The piecemeal solution is simply to add them one at a time. But suppose the numbers are larger:

$$127 + 128 + 129 + 130 + 131 = ?$$

The work of addition increases and one looks for a more general solution, especially if the series is a long one. The necessary recentering of the data is facilitated when this problem is presented:

$$18 + 19 + 20 + 21 + 22 = ?$$

And the problem becomes even less difficult, according to Wertheimer, when this problem is presented:

$$99.8 + 99.9 + 100 + 100.1 + 100.2 = ?$$

In general these problems used by Duncker and Wertheimer show that the difficulty of arithmetic problems is often due to the embedding of the essential relations within an initially confusing pattern. Duncker has shown that aids which help to reorganize the pattern into an easier pattern do facilitate solution, while other aids do not. Wertheimer's illustrations are intended to show that some ways of writing the numerical patterns indicate the essential relations more clearly than others.

5. Another source of difficulty is the sequence of problems. Some sequences build task-induced sets or habits of problem solving that are helpful in solving the later problems, while other sequences build up interfering sets. The preceding chapter discussed the effects of set on production of simple responses; the effect on the solution of complex problems is probably similar, but is complicated by the other difficulties described above.

SPATIAL AND MECHANICAL PROBLEMS

Favorite problems of the psychological laboratory are those that require the subject to go from one place to another, to manipulate some object or tool to attain a goal, to put things together, or to find a hidden object. It is quite proper that such problems should be emphasized in the laboratory because they occur frequently outside the laboratory also, in the lives of all mobile animals with manipulatory equipment.

THE MAZE PROBLEM. Good evidence on the mastery of spatial patterns comes from maze experiments with human subjects (see Woodworth, *1938;* Underwood, *1949*). Analysis of errors, of reproductions of the maze pattern by drawings, and of introspections discloses in general terms how the maze pattern is developed. The general outline of the maze pattern, at least the orientation of the goal from the start, is learned first, followed by filling in of the details. Learning

extends out from the entrance and from the goal. In a long maze there often seem to be intermediate points from which the maze is organized in both directions.

The usual maze problem is not hard to solve because each choice point presents only two alternative paths. Anyone with persistence will solve it eventually. The question of present interest is one of the processes by which the subject constructs an instrumental or cognitive pattern of the true path so he can go to the goal easily. Here the comments of the subjects are valuable because many report that they use nonspatial aids, frequently verbal aids. Comparisons of the subjects' reports of their methods with objective data on time and errors show that verbal aids are quite helpful (Warden, *1924;* Husband, *1931*). The verbal aid usually consists of a pattern of turns memorized verbally: "First is right, next left; watch for crook; next two right; go way back; etc." Naturally verbal aids are most useful for mazes in which the spatial relations are confusing, e.g., linear mazes. Outside the laboratory, when someone prepares for a trip into unfamiliar territory, he usually takes along a map or at least some verbal directions. These aids may be called instrumental patterns. They do not represent the environment completely; they are patterns of the pertinent aspects of the environment, pertinent, that is, to the immediate problem of getting from starting place to goal.

String-Pulling Problem. The process of analyzing a perceived situation to find out what leads to what is nicely illustrated by the

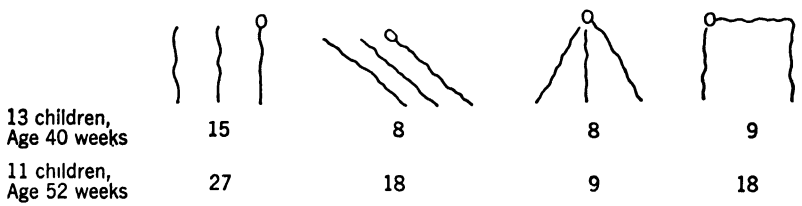


Figure 8.2. String-pulling problems. Showing percentages of correct first choices by young children. Extracted from Richardson (*1932*).

string-pulling problem. The subject is presented a pattern of strings and lure, his task being to pull the string that brings the reward. These problems are too easy for adult humans, who can perceive or "read off" the means-end relations at first glance, but are rather difficult for monkeys and young children. Figure 8.2 shows some string

problems used by Richardson (*1932*), with the percentages of correct solutions on the first try. According to the experimenter's observations, the children's problem-solving activities were determined by (1) the visual pattern presented at the moment, with the lure considered as the dominating element; (2) preëxisting tendencies to reach in a given direction or with a given hand; and (3) more general disturbing factors such as emotionality.

GEOMETRICAL PROBLEMS. Problems of plane geometry are very convenient for demonstrating reorganization of perceived patterns and have been frequently employed for this purpose by Gestalt psychologists. Little empirical evidence is available, but the important principles may be illustrated by summarizing Wertheimer's (*1945*) hypothesis of the difficulty involved in the problem of the vertical angles. The task is to prove that angles *a* and *c* of Figure 8.3 are equal. One can proceed algebraically, as follows:

$$a + b = 180°$$
$$c + b = 180°$$

Subtracting $a - c = 0$

Therefore $a = c$

According to Wertheimer this demonstration is hard to follow. Students may learn it mechanically, but the occurrence of mechanical errors in repeating the proof shows that the proof is not genuinely understood. The principal reason for the difficulty is that it is hard to perceive angle *b* as part of $a + b$, which equals 180°. It is the structure that must be grasped, and the relations of the parts to the structure, rather than the mastery of any particular equation or sequence of equations. Devices that aid analysis of the structure and show how angle *b* is contained in two larger patterns, as in Figures 8.4 and 8.5, should be helpful.

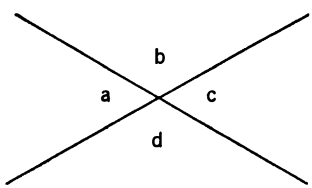


FIGURE 8.3.

Wertheimer's hypothesis of the difficulty in problems of this type could be checked experimentally. Aids that are constructed according to the hypothesis could be given to students along with aids based on some other hypothesis. As an alternative hypothesis one might include the difficulty of viewing a straight line as an angle of 180°. For many

students the concept of angle is limited to acute angles. A straight line is a straight line, a good figure, not an angle. Practice in dealing with straight lines as special kinds of angles might be an aid in this problem.

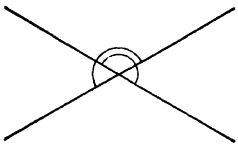


FIGURE 8.4.

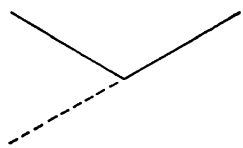


FIGURE 8.5.

CONSTRUCTION PROBLEMS. Construction puzzles have the advantage that part of the problem-solving activity consists of easily observed movements. Durkin's observations of college students' efforts to put flat pieces together to make a cross, for example, permitted her to describe their activities under the headings of trial and error, sudden reorganization, gradual analysis (see page 31).

If the puzzle permits a variety of free movements, unitizing and recording the movements is difficult. Cook (*1937*) used a disk-transfer puzzle that is solved by taking disks off one peg and putting them on others. Each such move can be treated as a unit of activity. Cook had his subjects, who were college students, about 30 in each group, work puzzles of two, three, and four disks. Each had to keep working until he discovered how to work each puzzle in the minimum number of moves. Records were kept of the number of moves made before discovering the best solution, and of the time required. The results are shown in the table below.

Number of disks	2	3	6
Number of moves in best solution	3	7	15
Average number of moves to solve	3.5	24.3	182.7
Average time in seconds	7.0	93.5	557.0

If we take the minimum number of moves as a measure of the complexity of each task, and the number of moves used in solving as a measure of difficulty, it is clear, for this kind of problem as for many others, that difficulty is determined in part by the complexity of the problem situation.

The complexity of the task seems to affect not only the difficulty but also the optimum distribution of practice periods. When Ericksen (*1942*) had college students work a maze and a problem box, distributed practice was better for the maze while massed practice was better for the problem box. As in the disk-transfer problem above, an indication of the complexity of each task may be obtained from the variability of the problem-solving activities. Average number of movements on the first trial of the maze was 14, declining to 10 on the last trial. For the problem box the average number of movements on the first trial was 78, declining to 6 on the last trial.

This way of measuring the complexity of a problem in terms of the variability of the behavior it evokes gives some objective meaning to the dimensional analysis mentioned in Chapter 2. Another possibility along this same line is to make up an artificial problem of discrete moves in such a way that modern information theory can be applied (Whitfield, *1951*).

Maier has carried out a series of studies of construction problems which show the effects of verbal instructions and other aids in facilitating solution. In one of Maier's (*1930*) experiments the subjects were faced with a rather difficult problem, constructing supports to hold two pendulums that would make chalk marks on the floor. The solution of this problem requires the use of several principles, or techniques, or parts of the complete solution pattern, e.g., wedging, clamping. All of these parts of the solution must be properly combined in the correct organization. Different groups of subjects were given (1) the problem only, (2) demonstration of the necessary parts of the solution, (3) instructions to combine the demonstrated parts of the solution, (4) a hint or suggestion about hanging a pendulum on a nail in the ceiling, and (5) all of these aids. Only in the group that was given all aids was there an appreciable number of solutions. None of the aids was sufficient by itself. The results of this experiment do not favor any one aid over the other, but when the experiment was repeated by Weaver and Madden (*1949*) explanation of the parts of the solution seemed to be the most helpful aid.

In the second experiment on a similar construction problem, Maier (*1931*) made his hints more specific. Two strings hanging from the ceiling are to be tied together, though they are so far apart that the subject cannot reach both at once. Only 39 percent of the subjects got the "correct" solution, tying a weight on one string and setting it

swinging, in ten minutes. When the experimenter brushed one string to make it swing and handed a weight (pliers) to the subject, 38 percent more were able to solve the problem.

Varying the preparation for this two-string problem by changing the instructions has a definite effect on solution time (Battersby, Teuber, and Bender, *1953*). One group, told to use only the five objects placed in front of it, solved in a median time of two minutes. Another group, told to use any object in the room including those objects in front of it, solved in seven minutes. Median time for another group, told to use any object in the room but not told specifically about the objects that happened to be lying on the table, was 18 minutes. These differences due to instructions are large. The effects of brain damage on performance of this problem were, by contrast, quite small.

Verbal aids that operate in an indirect or mediate way are also useful. Before working this two-string problem, Cofer (*1951*) had his subjects learn various lists of words. One group that memorized rope, swing, and pendulum, among others, produced more pendulum solutions than those who memorized only irrelevant lists.

A third experiment by Maier (*1945*) carries the analysis further and attempts to evaluate the contribution of previous experience to problem solving. The problem was to build a hat rack from certain materials at hand. The solution was to clamp two boards together, and wedge the assembly between floor and ceiling so that the clamp in the middle could be used to hold a hat. Of a group of 25 subjects who tried the problem without aid, only 6 solved it. Each subject in another group worked together with the experimenter to solve a pendulum problem, the solution to which included the clamping and wedging principles. Then the pendulum structure was removed and these subjects attacked the hatrack problem. Almost half, 12 out of 25, of this group profited by experience and solved the problem. But the best results were achieved by a third group that had the initial practice with the pendulum problem and also was allowed to look at the completed pendulum support while working on the hatrack problem. Of this group of 25, 18 succeeded. Evidently some experience with the parts of a solution aids organization of the whole, but an example in view aids even more. (Other aspects of this experiment were described on page 122.)

These three experiments by Maier are concerned chiefly with the

relation between the parts of a solution, e.g., principles, tools, techniques, etc., and the whole. The parts may be present in view or previously used by the subject, yet not used in solution of the present problem. The thinker may be familiar with pendulums, for example, and he may have used pliers. But pliers have been used to grasp things with; their function is embedded in the instrumental pattern of grasping. A considerable reorganization of the concept of pliers is necessary before they will be used as a pendulum bob. The hints aid this reorganization.

Another, overlapping source of difficulty in these experiments is persistence of set. The peculiarities of the problem establish a task-induced set or direction, centering attention on some objects rather than others. Part of the contribution of the specific verbal instructions is to help the thinker change his set. It is possible also that general instructions about changing set may be helpful. Maier (*1933*) tried out his hypothesis by giving an experimental group a 20-minute lecture on problem solving with special emphasis on the necessity of shifting set. This group did better in solving problems immediately afterward than a control group not so instructed. In the preceding chapter it was pointed out that Luchins (see page 199) obtained similar small improvement by telling his subjects not to be blind at the appropriate time.

The Gottschaldt figures (see page 105) illustrate how a pattern may be concealed in another inclusive pattern. Duncker (*1945*) has described several experiments that illustrate how an object can be concealed, or at least made unavailable for problem solving, by being included as part of an irrelevant pattern. His box problem may be taken as an example. The problem is to support three candles on the door at eye level, ostensibly for an experiment on vision. On a table lie many objects, including a few tacks, and three small boxes of different shapes and colors. The correct solution is to tack the boxes on the door as platforms for the candles.

The experimental variable in Duncker's experiments is preutilization, or embedding. For an experimental group of seven subjects, the three boxes were filled with candles, tacks, and matches, respectively. For the control group of seven subjects, they were empty. The hypothesis is that perceiving the boxes used as containers precludes using them as platforms. Only three out of seven of Duncker's experimental subjects, who saw filled boxes, were able to solve the problem; but all

seven control subjects, who saw empty boxes, were successful. Similar results were obtained with several other problems of this kind.

Birch and Rabinowitz (*1951*) have criticized Duncker's experiments on the grounds that the critical preutilized solution and the alternative solution possibilities were not equated. (Furthermore, some of the subjects apparently were assigned to experimental and control groups *ex post facto*. See note to Duncker's Table 3.) So Birch and Rabinowitz carried out a similar experiment with a small number of subjects but with good controls. They employed Maier's two-string pendulum problem, which requires the use of an object as a pendulum bob.

In this experiment a relay and a switch were on a table in front of the subject. A control group of six subjects approached the problem with no pretraining and, as expected, three used the relay as a pendulum bob and three used the switch. One experimental group of ten subjects had a pretest task of constructing a circuit, using the relay. When they later attacked the pendulum problem, they all used the switch. Another experimental group of nine subjects had a pretest task of constructing a circuit, using the switch. When they later attacked the pendulum problem, all but two used the relay. Thus 17 of the 19 experimental subjects avoided the object that they had used for a different purpose and made use of the object not previously used. This difference is statistically significant, confirming Duncker's conclusions. Confirmation has been added by Adamson (*1952*) also.

CONCLUSIONS THUS FAR

A few general principles can be gleaned from a review of the evidence assembled above. In the first place it is obvious that there is no one process or sequence of process by which all problems are solved. If we examine a wide variety of problems, we see that they vary greatly in the material of the problem, in the chief causes of difficulty, and in the kinds of aid used in solving them. Burack (*1950*) studied only three problems, but found considerable variation from one to another in kinds of attack. We can assume, in a very general way, that all problem solving consists of (1) survey of the problem and analysis of the goal, (2) production of promising solution attempts, and (3) judgment or evaluation of the attempts. But if the problem is clearly posed, hardly any analysis is necessary. And in some problems the essential intellectual activity is not one of continuous production, as in the tasks discussed in the preceding chapter, but one of brief, perhaps

sudden, reorganization of the patterns at hand. Furthermore, if production is aided by a specific search model, the only solution produced may be obviously correct, hence no separate act of evaluation is necessary. Perhaps one can demonstrate logically that certain steps are always required, but if such steps cannot be observed or reliably inferred, such a logical demonstration is psychologically irrelevant.

For convenience, experiments on the production of patterns and problem solving have been grouped roughly according to the material or content of the problem, namely, verbal, numerical, spatial. But a backward glance over these pages will permit an interesting observation: a problem constituted of any kind of material may be solved with the aid of instrumental patterns of any other kind of material. Spatial problems, e.g., mazes, are often solved with the aid of verbal or numerical patterns. Geometrical patterns may be solved with the aid of numerical and verbal aids, i.e., analytical geometry. Numerical problems are often solved by stating and restating the numerical relations in words or by diagramming the relations spatially. Verbal problems may be solved by quantifying the data or by the aid of diagrams, e.g., by Euler's circles. The kind of instrumental patterns used in this way probably depends on personal preferences, previous training, the nature of the problem, and so on, though this question has not been attacked empirically. The implications of this generalization for the analysis of individual differences by factor analysis will be discussed later.

When we look for generalizations about the sources of difficulty in problem solving, the evidence of the preceding pages is reasonably satisfactory. Several experiments have demonstrated the importance of (1) novelty or lack of familiarity of the material of the problem, (2) the abstractness of the material, (3) the complexity of the problem situation, (4) embedding or concealment of a critical aspect of the solution in an inclusive pattern, and (5) persistence of set. The first four causes of difficulty have been conveniently illustrated and summarized by Welch (*1947b*).

CREATIVE THOUGHT

Problems that come under the heading of creative thought, such as writing a poem, inventing a labor-saving device, or designing a psychological experiment, are not entirely different from the more prosaic varieties discussed above. The preparation, as we have seen in

Chapter 6, is less definite at the outset and the production process necessarily becomes irregular as the preparation changes. Much thought in fact may be given over to the production of alternative preparations.

If one is serious about designing a fireplace, for example, he may spend some time considering how to begin, rehearsing what he wants in a fireplace, what he can afford to spend, remembering the possibilities of error, etc., and from these assembling a search model to guide the subsequent productive flow of tentative designs. The first outcome, therefore, may be not a first draft of a design but a decision to sharpen a pencil or to visit a library. In other words the initial subproblem may be to find the best plan of action, and the production process guided by the first formulation of the subproblem may be such patterns of action as: work here, adjust the light, do most of it myself, design first and material second, use scale of two inches to the foot, and the like. There is nothing particularly new here, but as the thinker fluctuates from preparation for thought to emission of ideas, to criticism of these and back and forth again, one would expect considerable irregularity in the process. Nevertheless two very interesting phenomena of productive thought, interesting just because they are irregular, have been noticed for many years, e.g., by Ruger (*1910*), and named by Graham Wallas (*1926*) incubation and illumination.

INCUBATION

As Wallas describes it:

> The Incubation stage covers two different things, of which the first is the negative fact that during Incubation we do not voluntarily or consciously think on a particular problem, and the second is the positive fact that a series of unconscious and involuntary . . . mental events may take place during that period (p. 86).

Many thinkers have reported, often in colorful detail, something like incubation as thus described. (See Woodworth, *1938;* Hutchinson, *1949;* Ghiselin, *1952*). Such unsystematic reports are open to some criticism, to say the least. The most serious attempt to check on this phenomenon is that of Patrick, working with poets (*1935*) and with artists (*1937*). Preceding each investigation, she asked her subjects these questions, among others:

> When you get ready to write a poem (or to make a picture) do you

incubate it a while first? If so, describe the process of incubation to me. How long does it usually last? What are its characteristics?

Sample replies follow:

First I have a mood, then I get two or three lines. Sometimes I carry those two or three lines around and sometimes not. A sudden stimulus starts the incubation of a poem, whether a mood or a line. A mood may lie fallow from a half-hour to a year, and then it comes back.

If I have a feeling for a poem I would read something that would keep the suggestion going and carry some sort of a living relation to me. For instance, when I wrote the poem "Morning Glories" the mood, as a sort of stillness, lasted for some time, and after a while the poem came to me. In this way ideas incubate.

These comments show that the poets and artists were familiar with the idea of incubation and able to describe it. When Rossman (*1931*) and Platt and Baker (*1931*) sent questionnaires to numerous scientists and inventors, it was clear from their replies also that they often lay their problems aside for a time, after preparing for them, to return to them later.

To nail this phenomenon down with somewhat more objectivity Patrick recorded the comments of her subjects as they worked (see page 29) and "noted those cases in which the idea of the poem appeared earlier in the report, after which the subject talked of various things, and then this original idea reappeared as the subject of the poem. This is the fact in over two-thirds of the cases." For the artists the proportion was four-fifths.

In both of these studies Patrick compared experts with novices and found that, while they differed in respect to techniques used and the merit of the finished product, they did not differ in respect to the stages of thought. Artists are typically more systematic in their methods of work than nonartists and display more control over their creative activities (Eindhoven and Vinacke, *1952*).

Interferences resulting from emotional and other causes may drop out if the thinker abandons his problem for a time and then returns to it. For this reason one would expect that incubation would be common in thinking about one's own problems. But such phenomena are not often reported, at least not by that name, perhaps merely because no one has bothered to ask the right questions.

ILLUMINATION

Graham Wallas wrote: ". . . the appearance of the 'happy idea' together with the psychological events which immediately preceded and accompanied that appearance, I shall call Illumination" (*1926*). Others have spoken of the illumination as an insight or a flash experience or a downpouring of Heaven's blue.

Most active thinkers who are willing to trust their own introspection, as well as those who are willing to trust the introspection of others, will agree that something of this sort does occur. But how frequently and under what conditions it occurs, no one can say with any conviction because the evidence is introspective and anecdotal. These reports often mention, as would be expected, that previous preparation is a prerequisite for illumination, but that is hardly a novel discovery. The most specific point that may be gleaned from study of these anecdotes is that, after the preliminary labor of loading up with the facts and turning the problem round on all sides, the production of happy thoughts seems to be favored by walking. If this agreement is anything more than an expression of what thinkers of the past consider the good life, it may indicate something about an optimum degree of muscle tension.

E. D. Hutchinson (*1949*) reviewed more or less systematically many comments of eminent thinkers of the past and present about their most illuminating insights. He stated his generalizations as follows:

> Insights, aside from those which appear while one is deliberately engaged upon the problem, come most frequently, I believe, under three preferred conditions: during or just after periods of rest and relaxation; in periods characterized by a slight mental abstraction or dissociation which in itself furnishes a momentary relaxation; during periods of light physical activity, usually of a more or less repetitive and automatic character, which give relief from the insistent tensions involved (p. 120).

Patrick (*1935*) gave the term illumination a rather general meaning, neglecting the criterion of suddenness. In her study of poets she recorded thought changes as marking the period of preliminary preparation for creative thought and "lines as first drafted" as evidence for illumination. In her study of artists (*1937*) she used "general shapes of objects that were drawn for the first time." In both cases illumination as thus conceived was most frequent during the second and third quarters of a creative effort.

Perhaps these phenomena of incubation and illumination have received more discussion than they deserve. That thinking requires time and effort is not unexpected. That the central idea, when it comes, often comes suddenly, with a flash, is interesting but not startling unless one assumed that it should come gradually. That an idea recurs from time to time, unnoticed meanwhile, is a matter of more interest, but not startling either, unless one expected thought to be regular and attention continuous. Incubation and illumination are fascinating and perhaps mysterious only if one expects that the process of productive thought is always open to subjective observation. If that assumption had not been made, its disproof would not have been necessary.

EVALUATION AND REVISION

Creative thinkers and students of creative thought have frequently noted that creative production is usually followed by critical evaluation or judgment of the creative output. The evaluation or judgment process is hard to observe, of course, but an attempt at a general account of judgment will be made in the next two chapters. It is safe to say that the productive process is usually followed by a two-category judgment: "It is good enough. I'll stop." Or, "It is not good enough. I must continue." If the thinker continues, he may begin all over again with a radical reorganization, or he may merely revise what he has done.

Revision is relatively easy to observe. Poets strike out one word and put in another. Artists change details or add background without altering the central theme. Inventors look for sources of friction or minor difficulties. Patrick (*1935, 1937*) noted that her poets and artists did most of their revising during the last quarter of a creative enterprise.

Such revision, elaboration, and polishing is, of course, a minor productive activity. There is no reason to suspect that the process differs in any essential way from the more comprehensive productive activities discussed above.

CONDITIONS OF CREATIVITY

More important than the subjective phenomena of creative thought are the conditions that facilitate and impede accomplishment. It is very difficult to control the conditions of creative thought, as the conditions of laboratory problem solving have been controlled, hence

convincing evidence is lacking. But some plausible inferences may be derived from laboratory information on the conditions of fluency and flexibility and the organization of patterns.

There can be little doubt that creative accomplishment, like other intellectual assignments, requires a certain minimum of practice or familiarity with the material of the problem. The preparation process was described in the preceding chapter. It is merely necessary to add that the first published poem or novel is usually preceded by one or more rejected manuscripts. Invention of a new bookkeeping system is usually preceded by familiarity with existing bookkeeping systems, and the development of a new germicide is usually dependent upon considerable knowledge of the advantages and disadvantages of existing germicides. It was not Buddha but Moses who saw God in the burning bush.

Especially when the situation is complex, additional familiarity with relationships within the situation seems to be necessary. It requires considerable familiarity with the operation of a tax structure, for example, under different conditions before one could invent a safe way of reducing one's payments or devise a more foolproof system.

But the relation between experience and creativity is certainly not simple. Some kinds of practice conditions, as noted in the preceding chapter, lead to a rigid or routine solution which by definition cannot be creative. Perhaps it is practice that emphasizes the procedures rather than the goals, the means rather than the ends that inhibits creative accomplishment; but no pertinent information is available. Another possibility is that practice in which the goal is emphasized by frequent changes would be more productive, or practice in breaking up patterns embedded in larger patterns.

Many creative thinkers have recommended (see Hutchinson, *1949*) what may be called an alternation technique. They collect and organize facts, requirements, hunches, etc., pushing the solution as far along as they can, then deliberately abandon the problem for a while. When the problem is picked up again, interferences may have dropped out, sets may have changed, and relationships between the facts may look different.

We have seen that fluency and flexibility are reduced by emotional disturbances, so it would be plausible to suppose that creative thought is thus inhibited. Barker, Dembo, and Lewin (*1941*) have shown that the constructiveness of children's play is reduced by frustration. Even

this principle is complicated, however, because the emotion may have the effect of increasing persistence and thus overcoming its inhibitory effects. We must not confuse the deleterious effects of a brief emotional outburst with the facilitative effects of long-continued motivation. If any thinker continues to work on a problem, because of either extrinsic or intrinsic motivation, for weeks and months, his chances of creative accomplishment, especially through incubation, are greatly increased. In fact the occurrence of incubation implies that the motivation persists over an appreciable period of time.

Some writers have claimed that their flow of ideas is improved by alcohol and other intoxicants. Experiments on fluency of production, reviewed in the preceding chapter, do not support this assertion in general, but it is conceivable that under some conditions release of inhibition by drugs could improve the quantity if not the quality of productive thought. As Ghiselin (*1952*) put it, the activities provoked by drugs and alcohol are hallucinatory rather than illuminating. Certainly, the final evaluation and revision of any rhapsodically produced ideas are best carried on under cool, sober conditions.

INDIVIDUAL DIFFERENCES

Conventional tests of intellectual abilities are based on individual differences in many of the activities described in this chapter. Years of correlational analysis have disclosed the general structure of relationships between such tests. It is safe to say, as a broad general principle, that success in solving one kind of problem correlates to some degree with success in solving other kinds. Success in solving many kinds is approximately what is meant by general ability, or general intelligence —though the score on any particular test of general intelligence will be influenced somewhat by the relative weights of the various kinds of problems in the composite.

In addition to this general rule of correlation between abilities, clusters or groups of abilities can be discovered. Tests that require various operations on verbal materials correlate highly with other verbal tests, but not so highly with tests that involve numerical relations or spatial relations (Thurstone, *1938;* Wolfle, *1940;* Vernon, *1950*). At least the grouping of abilities is clear enough to justify the organization of part of this chapter around such groups.

A later chapter will be devoted entirely to individual differences. The question for the remainder of this chapter is: What can we add

to this general structure of abilities? More specifically, do the experiments described in this chapter disclose any new intellectual traits not ordinarily included in the study of individual differences? We may hope to clarify and broaden the psychology of individual differences and at the same time to further the analysis of the thought processes.

ABSTRACT VERBAL MATERIAL

Tests of anagrams, artificial language, vocabulary, verbal similarities and differences, reading, syllogisms, and the like have frequently been used to get scores for individual differences. All these tests involving verbal materials are correlated, and the general factor in these tests has become known as the verbal factor, or verbal ability. Naturally, success on any of these problems will correlate with general intelligence, in which the verbal factor is well represented, but will correlate more closely with a good test of the verbal factor. The more abstract tests, e.g., syllogisms, abstract concepts, difficult verbal relations, have the highest correlations with general intelligence. A good case can be made for the argument that ability to solve verbal problems, or to organize abstract verbal patterns, is the most important component of general intelligence.

QUALITATIVE DIFFERENCES IN CONCEPTUALIZATION. Many attempts have been made to separate people into categories or types in respect to abstract thinking, generalizing, kinds of concepts and principles used, and other signs of conceptualization. A free-response method is employed, and the test material includes incomplete sentences, questions about cause and effect, proverbs to be interpreted, words to be compared and contrasted, and words or objects to be put into groups. The results of these experiments should be interpreted in relation to two well-known factors: (1) verbal ability and (2) set.

1. Individual differences in verbal ability can be tested easily and reliably. Relations between verbal ability and many other variables, e.g., numerical ability, age, social class, and education, are fairly well known. There is little to be gained by inventing *ad hoc* variables or classification schemes if a well-known continuous variable accounts approximately as well for the results. Results of tests of concept formation and verbal problem solving should be compared with scores for verbal ability rather than, or in addition to, scores for general intelligence because tests of verbal ability are more homogeneous and

would be more closely related to performance on any task involving words.

2. Set. Any conversation, interview, or psychological test takes place in a social situation, which involves interpersonal relations between investigator and subject. Often it can be assumed, as in experiments with college students, that those relations are the same for all subjects; that each is prepared in approximately the same way by the experimenter's instructions. But this assumption cannot be made without danger when children are compared with adults, when schizophrenic subjects are compared with normal subjects, or when subjects of the investigator's culture are compared with subjects of a strange culture. It is more than a matter of willingness to coöperate; it includes subtle social factors of great importance for conceptualization, such as sensitivity to social and self-criticism and readiness to conform to the norms of rational communication. In experiments with children and with psychiatric patients, one cannot freely assume that the subject can maintain a set long enough to perform the required intellectual operations and report them to the investigator. Individual differences in set have been discussed in earlier chapters.

Many attempts have been made to compare children, and children and adults, in respect to the concepts and principles they use in explaining physical and social phenomena (Werner, *1940;* Vinacke, *1951*). The early attempts, e.g., Piagêt (*1929, 1930*), used a simple interview method. Children were asked questions like "What makes the clouds move," and the answers were classified in such categories as animistic and realistic. More recent studies (Deutsche, *1937;* Russell and Dennis, *1939*) have improved the technique, refining the classification procedures and using two or more classifiers.

The most general statement that can be made of the results of these studies is that, as children grow and learn, their conceptualizing becomes more and more like that of the adults they learn from. At most age levels, including adulthood, all varieties of concepts and principles are used by at least some individuals. As one would expect, maturity of conceptualization correlates with verbal ability and age (Deutsche, *1937;* Ordan, *1945*). The great importance of education as compared with other variables may be seen in Meltzer's (*1925*) results of a study of children's social concepts. For 333 children in Grades 4 to 12, maturity of conceptualization correlated .36 with occupational status of

parents, .55 with age, .58 with mental age, .69 with grade, and .80 with general educational achievement.

This trend from childish to adult forms of thought, whether continuous or saltatory, is part of a massive socialization process, for the child gets his examples and his rewards from social interaction. Children are limited not only by their verbal abilities but also by their experience, information, and range of interests. Thus they can be called egocentric (Piaget, *1926;* McCarthy, *1946*). As the child lives and learns in the usual give and take with other people, he finds it easier to adopt another's point of view and to conform to the norms of rational communication. Everyone organizes his world with himself at the center, but the normal adult under favorable conditions can reorganize the pattern so as to examine himself and the world from other angles.

The statement has occasionally been made that primitive people think more animistically or more magically or more concretely than people of industrialized cultures. But many instances of naturalistic and abstract thinking can be collected from primitive philosophers if one looks for them (Radin, *1927*). The evidence on both sides is only anecdotal. There may be differences between any two cultures in frequency of use of various kinds of concepts and principles, but no one has yet compared representative samples in this respect.

Many attempts have been made also to find qualitative differences between normal and psychiatric subjects in respect to conceptualization or abstract thinking. The bizarre language of schizophrenic patients has attracted particular attention. A thorough survey of the evidence by Hunt and Cofer (*1944*) indicates that the pathology of thought in schizophrenic patients is largely one of difficulty in maintaining a set. Wegrocki (*1940*), for example, started out to study generalizing ability in schizophrenia, using tests of verbal analogies, essential differences in words, and interpretation of proverbs. He found the expected overall inferiority of schizophrenics as compared to children and normal adults, and grouped the errors of the schizophrenics in the following categories: domination by a particular set, inability to organize the stream of associations, tendency to exhibit aphasic phenomena, lack of ability to focus on a problem situation, and periphrastic approximation. An example of periphrastic approximation was given by one of Wegrocki's patients when asked to complete the analogy: "clothes are to tear as dishes are to ———."

The patient said, "If clothes tear then dishes *drop* and break," but by the time he started to write, he substituted *"slopes"* because "that's the way *a line drops*." Difficulties in maintaining a set, which probably account for much of the psychopathology attributed to difficulty in abstract thinking, were considered in the preceding chapter.

A promising free-response technique for describing cognitive patterns has recently been devised by Rokeach (*1951a*). The subject is shown a list of ten terms: Buddhism, Capitalism, Catholicism, Christianity, Communism, Democracy, Fascism, Judaism, Protestantism, Socialism, and asked to describe in what way these terms are related. If the replies give evidence that all terms are organized in a broad, integrated structure, the pattern is called "comprehensive," e.g., "all are beliefs." If all terms are taken into consideration but organized into two or more apparently unrelated substructures, the pattern is called "isolated," e.g., "five are religions, five are governments." And if some items are related but others simply neglected, the organization is called "narrow."

When 144 college students were asked to perform this task, 66 of their papers were classified as "comprehensive," 59 as "isolated," and 19 as "narrow." Classifications by a second judge indicated good reliability. The three groups are significantly different in general intelligence, the "comprehensive" subjects getting the highest scores, the "narrow-minded" the lowest (Rokeach, *1951b*). When these 144 students were given a printed test of ethnocentrism, the "comprehensive" subjects were least ethnocentric and the "narrow-minded" subjects most ethnocentric. Since comprehensiveness and ethnocentrism as thus measured are both related to general intelligence, and would probably be more closely related to verbal ability, we cannot yet be sure whether this testing technique yields more information that is obtained from well-known tests. Rokeach believes that qualitative description of cognitive patterns as "comprehensive," "isolated," and "narrow" is more informative than a quantitative score on a standard test.

NUMERICAL MATERIAL

Individual differences in forming numerical patterns have not been studied as thoroughly as verbal patterns, perhaps because the material is more homogeneous, but the evidence permits a few generalizations.

At all age levels success in solving arithmetic problems correlates with success on tests of general intelligence. Tests of simple arithmetic

operations, such as addition and subtraction, can be grouped together to form a numerical group ability or factor that is relatively distinct from verbal and spatial abilities. Tests of arithmetic problem solving have more in common with tests of verbal ability, such as a reading test. As in the case of verbal material, the more abstract numerical tasks are better measures of general ability than the simple specific tasks. Certainly scores on such tests are potentially useful predictors of success in solving numerical problems.

SPATIAL AND MECHANICAL MATERIAL

Tests of individual differences in perceiving, comparing, organizing, breaking up, and manipulating perceived spatial patterns are well known. Scores on any of these tests correlate more or less with general intelligence, but they have more in common with scores on other spatial tests. The group factor in a variety of such tests has become known as the space factor and, because of its closer relation to mechanical than to academic assignments, as practical intelligence. As in the case of verbal and numerical tests, the more abstract spatial tests that require visualization and reorganization, e.g., counting blocks, some of which are not seen but inferred, are better measures of general ability than tests of simple perceptual skills.

Curiously, tests of spatial ability have seldom been used in laboratory experiments on the solution of spatial and mechanical problems. Harrell (*1940*) has shown, however, by a factor analysis of several mechanical ability tests, that the space factor, composed of printed tests of spatial ability, is the principal factor in ability to construct useful mechanical objects and tools. A visualization factor may also be identified in military samples (Guilford, *1947*). It may turn out that this space factor is an important contributor to success on the geometrical and construction problems described above. Certainly one should be cautious about identifying a new problem-solving ability in construction puzzles unless he has compared it with well-known, easily measured abilities.

A promising method for qualitative analysis of children's solutions to spatial problems has been described by Werner (*1944*). Two square boards are used, with holes for marbles. The examiner exposes a pattern of marbles on one board and asks the child to reproduce it on the other. A record of the child's performance is kept on an appropriate blank.

In a sample of mentally retarded children, those of higher intelligence do these problems quantitatively better than the more retarded, of course, but qualitative differences also may be observed. Regardless of success, the children's performance may be classified, with unknown reliability, as continuous, constructive, and irregular. The continuous procedure consists of placing one marble, then one next to it, one next to that, and so on. The constructive procedure is to break up the pattern into squares, triangles, and other substructures, and form these substructures into the whole. The irregular procedure is neither of these, but a hit-and-miss approach. With increasing mental age, the irregular and continuous methods are used less frequently, the constructive approach more frequently. Werner has used this kind of qualitative analysis also in differentiating endogenous and exogenous types of mental deficiency.

ABSTRACTION AND FLEXIBILITY

For convenience of exposition, flexibility in shifting set was discussed in the preceding chapter, while this chapter has considered the organization and reorganization of abstract patterns. But, as pointed out earlier, ability to abstract any feature of a problem situation from its surroundings and manipulate it in any way requires ability to maintain a set for that feature and shift the set if necessary. Is there any correlation, then, between ability to shift set and ability to handle abstract material?

Luchins (*1951*) has studied this relationship between set and abstraction in a sample of 91 college women. As a measure of abstract thinking, he used the similarities test of the Wechsler-Bellevue Scale of Intelligence. This test asks for similarities between two words, e.g., eye and ear, each response to which can be classified as abstract, concrete-functional, or failure. It turned out that the 25 subjects who were most flexible on the volume-measurement problems gave more abstract similarities, while the 22 who were most rigid produced more concrete-functional similarities and more failures.

Luchins' data may also be analyzed in the conventional way, failures being given a score of 0, concrete-functional similarities a score of 1, and abstract similarities a score of 2, in order to compare them with other data. By this procedure the most flexible group gets the highest average score for abstract thinking, and the most rigid group the lowest average score. Two smaller groups of medium rigidity get

medium scores for abstract thinking. The relationship is small, however, probably because the Shipley test requires some flexibility in shifting set, while the similarities test is a fairly pure measure of ability to use abstract concepts. These data confirm the low positive correlations between flexibility on the volume-measurement problems and verbal ability mentioned in the preceding chapter (see page 214).

If we assume that success on Maier's two-string problem (see page 256) depends on breaking up and reorganizing a perceptual pattern, as Maier does, some correlations recently reported by Battersby, Teuber, and Bender (*1953*) are interesting. Scores on the two-string problem, in a sample of 20, correlate .72 with scores on the Gottschaldt figures, which also require reorganization of a perceived pattern. These scores correlate only .46 with scores on a variation of the Weigl sorting test, which is presumably a test of flexibility in shifting set.

Aligning the evidence from these comparisons of flexibility in shifting set and flexibility in manipulating abstract patterns, together with some evidence from the preceding chapter, permits two generalizations. First, there must be some relationship between the two abilities. The relationship will be lower in a college sample, of course, than in a sample with a broader range of ability. Second, flexibility in shifting set has only a low relationship to pure abstraction, as indicated by measures of ability to generalize, ability to use abstract concepts, and conventional tests of verbal ability; but it is more closely related to tests that require breaking up perceptual or cognitive patterns, or shifting from one kind of abstraction to another.[1]

REFERENCES

Ach, N., 1921. *Über die Begriffsbildung,* Untersuchungen zur Psychologie und Philosophie, Bamberg, C. C. Buchner.

Adamson, R. E., 1952. Functional fixedness as related to problem solving: a repetition of three experiments, *J. Exper. Psychol., 44:*288–291.

American Institute of Public Opinion, 1939. News release of July 9.

Anderson, G. L., 1949. Quantitative thinking as developed under connectionist and field theories of learning, in *Learning theory in school situations,* by E. J. Swenson, G. L. Anderson, and C. L. Stacey, Minneapolis, University of Minnesota Press.

Asch, S. E., 1946. Forming impressions of personality, *J. Abnorm. Social Psychol., 41:*258–290.

[1] Individual differences in creative thinking could logically be discussed at this point. It is more convenient, however, to consider this topic as part of a later general discussion of individual differences, following an examination of research on judgment.

Barker, R. G., Dembo, T., and Lewin, K., 1941. *Frustration and regression: an experiment with young children,* Iowa City, University of Iowa Studies in Child Welfare, 18, No. 1.

Battersby, W. S., Teuber, H. L., and Bender, M. B., 1953. Problem-solving behavior in men with frontal or occipital brain injuries, *J. Psychol., 35:* 329–351.

Birch, H. G., and Rabinowitz, H. S., 1951. The negative effect of previous experience on productive thinking, *J. Exper. Psychol., 41:*121–125.

Brotherton, D. A., Read, J. M., and Pratt, K. C., 1948. Indeterminate number groups, *J. Genet. Psychol., 73:*209–236.

Brownell, W. R., 1944. Rate, accuracy, and process in learning, *J. Educ. Psychol., 35:*321–337.

Brownell, W. R., Doty, R. A., and Rein, W. C., 1941. *Arithmetic in Grades I and II,* Durham, N. C., Duke University Res. Stud. in Educ., No. 6.

Brownell, W. R., and Stretch, L. B., 1931. *The effect of unfamiliar settings on problem-solving,* Durham, N.C., Duke University Res. Stud. in Educ., No. 1.

Burack, B., 1950. *The nature and efficacy of methods of attack on reasoning problems,* Psychol. Monogr., 64, No. 313.

Burt, C., 1919. The development of reasoning in school children, *J. Exper. Pedagogy, 5:*68–77, 121–127.

Buss, A. H., 1950. A study of concept formation as a function of reinforcement and stimulus generalization, *J. Exper. Psychol., 40:*494–503.

Cofer, C. N., 1951. Verbal behavior in relation to reasoning and values, in *Groups, leadership and men,* ed. H. Guetzkow, Pittsburgh, Carnegie Press.

Cook, T. W., 1937. Amount of material and difficulty of problem solving: II. The disc transfer problem, *J. Exper. Psychol., 20:*288–296.

Davidon, R. S., 1952. The effects of symbols, shift, and manipulation upon the number of concepts attained, *J. Exper. Psychol., 44:*70–79.

Deutsche, J. M., 1937. *The development of children's concepts of causal relations,* University of Minnesota Institute of Child Welfare Monograph Series, Minneapolis, University of Minnesota Press.

Devnich, G. E., 1937. Words as "Gestalten," *J. Exper. Psychol., 20:*297–300.

Dixon, J. C., 1949. Concept formation and emergence of contradictory relations, *J. Exper. Psychol., 39:*144–149.

Drever, J. I., 1934. The pre-insight period in learning, *Brit. J. Psychol., 25:*197–203.

Duncker, K., 1945. *On problem-solving* (trans. L. S. Lees), Psychol. Monogr., 58, No. 270.

Eindhoven, J. E., and Vinacke, W. E., 1952. Creative processes in painting, *J. Gen. Psychol., 47:*139–164.

English, H. B., 1922. An experimental study of certain initial phases of the process of abstraction, *Am. J. Psychol., 33*:305–350.

Ericksen, S. C., 1942. Variability of attack in massed and distributed practice, *J. Exper. Psychol., 31*:339–345.

Ewert, P. H., and Lambert, J. F., 1932. The effect of verbal instructions upon the formation of a concept, *J. Gen. Psychol., 6*:400–413.

Gengerelli, J. A., 1927. Mutual interference in the evolution of concepts, *Am. J. Psychol., 38*:639–646.

Ghiselin, B., 1952. *The creative process, a symposium,* Berkeley, University of California Press.

Gibbons, H., 1940. The ability of college freshmen to construct the meaning of a strange word from the context in which it appears, *J. Exper. Educ., 9*:29–33.

Grant, D. A., 1951. Perceptual versus analytical responses to the number concepts of a Weigl-type card sorting test, *J. Exper. Psychol., 41*:23–39.

Grant, D. A., and Curran, J. F., 1952. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards, *J. Exper. Psychol., 43*:408–413.

Guilford, J. P., 1947. Printed classification tests, Army Air Forces Aviation Psychology Program Research Reports, Report No. 5, Washington, D.C., Government Printing Office.

Harrell, W., 1940. A factor analysis of mechanical ability tests, *Psychometrica, 5*:17–33.

Heidbreder, E., 1924. *An experimental study of thinking,* Arch. Psychol., No. 73.

Heidbreder, E., 1946. The attainment of concepts: II. The problem, *J. Gen. Psychol., 35*:191–223.

Heidbreder, E., 1947. The attainment of concepts: III. The process, *J. Psychol., 24*:93–138.

Heidbreder, E., 1948. The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels, *J. Psychol., 26*:193–216.

Heidbreder, E., 1949. The attainment of concepts: VII. Conceptual achievements during card-sorting, *J. Psychol., 27*:3–39.

Heidbreder, E., Bensley, M. L., and Ivy, M., 1948. The attainment of concepts: IV. Regularities and levels, *J. Psychol., 25*:299–329.

Heidbreder, E., and Overstreet, P., 1948. The attainment of concepts: V. Critical features and contexts, *J. Psychol., 26*:45–69.

Hines, R., 1950. *The formation and retention of concepts as a function of their abstractness,* Hays, Kans., Fort Hays Kansas State College.

Hollingworth, H. L., 1935. The conditions of verbal configuration, *J. Exper. Psychol., 18*:299–306.

Hollingworth, H. L., 1938. Verbal Gestalt experiments with children, *J. Exper. Psychol., 23*:90–95.

Hovland, C. I., 1952. A "communication analysis" of concept learning, *Psychol. Rev., 59*:461–472.

Hovland, C. I., and Weiss, W., 1953. Transmission of information concerning concepts through positive and negative instances, *J. Exper. Psychol., 45*:175–182.

Hull, C. L., 1920. *Quantitative aspects of the evolution of concepts,* Psychol. Monogr., 28, No. 123.

Hunt, J. McV., and Cofer, C. N., 1944. Psychological deficit, in *Personality and the behavior disorders,* J. McV. Hunt (ed.), New York, Ronald Press.

Husband, R. W., 1931. Analysis of methods in human maze learning, *J. Genet. Psychol., 39*:258–278.

Hutchinson, E. D., 1949. *How to think creatively,* New York, Abingdon-Cokesbury Press.

Johnson, D. M., 1944. A modern account of problem solving, *Psychol. Bull., 41*:201–229.

Johnson, D. M., 1950. Problem solving and symbolic processes, *Ann. Rev. Psychol., 1*:297–310.

Kuo, Z. Y., 1923. A behavioristic experiment on inductive inference, *J. Exper. Psychol., 6*:247–293.

Long, L., and Welch, L., 1941. Reasoning ability in young children, *J. Psychol., 12*:21–44.

Long, L., and Welch, L., 1942. Influence of abstractness on reasoning ability, *J. Psychol., 13*:41–59.

Luchins, A. S., 1942. *Mechanization in problem solving,* Psychol. Monogr., 54, No. 248.

Luchins, A. S., 1951. The Einstellung test of rigidity: its relation to concreteness of thinking, *J. Cons. Psychology, 15*:303–310.

McCarthy, D., 1946. Language development in children, in *Manual of child psychology,* ed. by L. Carmichael, New York, John Wiley & Sons.

Maier, N. R. F., 1930. Reasoning in humans: I. On direction, *J. Comp. Psychol., 10*:115–143.

Maier, N. R. F., 1931. Reasoning in humans: II. The solution of a problem and its appearance in consciousness, *J. Comp. Psychol., 12*:181–194.

Maier, N. R. F., 1933. An aspect of human reasoning, *Brit. J. Psychol., 24*:144–155.

Maier, N. R. F., 1945. Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning, *J. Exper. Psychol., 35*:349–360.

Marks, M. R., and Ramond, C. K., 1951. A new technique for observing concept evocation, *J. Exper. Psychol., 42*:424–429.

Meltzer, H., 1925. *Children's social concepts,* Teachers College Contributions to Education, No. 192, New York, Columbia University.

Nissenson, M., and Sargent, S. S., 1941. Words as configurations, *J. Exper. Psychol., 28:*85–89.

Olander, H. T., 1931. Transfer of learning in simple addition and subtraction, *Elem. Sch. J., 31:*358–369, 427–437.

Ordan, H., 1945. *Social concepts and the child mind,* New York, King's Crown Press.

Oseas, L., and Underwood, B. J., 1952. Studies of distributed practice: V. Learning and retention of concepts, *J. Exper. Psychol., 43:*143–148.

Overman, J. R., 1931. *An experimental study of certain factors affecting transfer of training in arithmetic,* Educ. Psychol. Monogr., No. 29.

Patrick, C., 1935. *Creative thought in poets,* Arch. Psychol., No. 178.

Patrick, C., 1937. Creative thought in artists, *J. Psychol., 4:*35–73.

Piaget, J., 1926. *The language and thought of the child,* New York, Harcourt, Brace & Co.

Piaget, J., 1929. *The child's conception of the world,* New York, Harcourt, Brace & Co.

Piaget, J., 1930. *The child's conception of physical causality,* New York, Harcourt, Brace & Co.

Platt, W., and Baker, R., 1931. The relation of the scientific "hunch" to research, *J. Chem. Educ., 8:*1969–2002.

Radin, P., 1927. *Primitive man as philosopher,* New York, D. Appleton-Century Co.

Reed, H. B., 1946a. Factors influencing the learning and retention of concepts: I. The influence of set, *J. Exper. Psychol., 36:*71–87.

Reed, H. B., 1946b. The learning and retention of concepts: II. The influence of length of series; III. The origin of concepts, *J. Exper. Psychol., 36:*166–179.

Reed, H. B., 1950. The learning and retention of concepts: V. The influence of form of presentation, *J. Exper. Psychol., 40:*504–511.

Rees, H. J., and Israel, H. E., 1935. *An investigation of the establishment and operation of mental sets,* Psychol. Monogr., No. 210.

Richardson, H. M., 1932. *The growth of adaptive behavior in infants: an experimental study at seven age levels,* Genet. Psychol. Monogr., No. 12. 195–359.

Rokeach, M., 1951a. A method for studying individual differences in "narrow-mindedness," *J. Personality, 20:*219–223.

Rokeach, M., 1951b. "Narrow-mindedness" and personality, *J. Personality, 20:*234–251.

Rossman, J., 1931. *The psychology of the inventor,* Washington, D.C., Inventors Publishing Company.

Ruger, H. A., 1910. *The psychology of efficiency,* Arch. Psychol., No. 15.

Russell, R. W., and Dennis, W., 1939. Studies in animism: I. A standard-

ized procedure for the investigation of animism, *J. Genet. Psychol.*, *55*:389–400.

Sargent, S. S., 1940. *Thinking processes at various levels of difficulty,* Arch. Psychol., No. 249.

Sells, S. B., 1936. *The atmosphere effect: an experimental study of reasoning,* Arch. Psychol., No. 200.

Sells, S. B., and Koob, H. F., 1937. A classroom demonstration of "atmosphere effect" in reasoning, *J. Educ. Psychol.*, *28*:514–518.

Smoke, K. L., 1932. *An objective study of concept formation,* Psychol. Monogr., 42, No. 191.

Spence, K. W., 1937. The differential response in animals to stimuli varying in a single dimension, *Psychol. Rev.*, *44*:430–44.

Swenson, E. J., 1949. Organization and generalization as factors in learning, transfer, and retroactive inhibition, in *Learning theory in school situations,* by E. J. Swenson, G. L. Anderson, and C. L. Stacey, Minneapolis, University of Minnesota Press.

Thiele, C. L., 1938. *The contribution of generalization to the learning of the addition facts,* Bureau of Publications, Teachers College, No. 763, New York, Columbia University.

Thurstone, L. L., 1938. *Primary mental abilities,* Psychometric Monogr., No. 1, Chicago, University of Chicago Press.

Underwood, B. J., 1949. *Experimental psychology,* New York, Appleton-Century-Crofts.

Vernon, P. E., 1950. *The structure of human abilities,* New York, John Wiley & Sons.

Vinacke, W. E., 1951. The investigation of concept formation, *Psychol. Bull.*, *48*:1–31.

Wallas, G., 1926. *The art of thought,* New York, Harcourt, Brace & Co.

Warden, C. J., 1924. The relative economy of various modes of attack in the mastery of a stylus maze, *J. Exper. Psychol.*, *7*:243–275.

Weaver, H. E., and Madden, E. H., 1949. "Direction" in problem solving, *J. Psychol.*, *27*:331–345.

Wegrocki, H. J., 1940. *Generalizing ability in schizophrenia,* Arch. Psychol. No. 254.

Welch, L., 1940. The genetic development of the associational structures of abstract thinking, *J. Genet. Psychol.*, *56*:175–206.

Welch, L., 1947a. A behaviorist explanation of concept formation, *J. Genet. Psychol.*, *71*:201–222.

Welch, L., 1947b. The transition from simple to complex forms of learning, *J. Genet. Psychol.*, *71*:223–251.

Welch, L., and Long, L., 1940a. The higher structural phases of concept formation of children, *J. Psychol.*, *9*:59–95.

Welch, L., and Long, L., 1940b. A further investigation of the higher structural phases of concept formation, *J. Psychol., 10*:211–220.

Wenzel, B. M., and Flurry, C., 1948. The sequential order of concept attainment, *J. Exper. Psychol., 38*:547–557.

Werner, H., 1940. *Comparative psychology of mental development,* New York, Harper & Brothers.

Werner, H., 1944. Development of visuo-motor performance on the marble-board test in mentally retarded children, *J. genet. Psychol., 64*:269–279.

Werner, H., and Kaplan, E., 1950. The acquisition of word meanings: A developmental study, *Society for Research in Child Development, 15*:No. 51.

Wertheimer, M., 1945. *Productive thinking,* New York, Harper & Brothers.

Whitfield, J. W., 1951. An experiment in problem solving, *Quart. J. Exper. Psychol., 3*:184–197.

Wilkins, M. C., 1928. *The effect of changed material on ability to do formal syllogistic reasoning,* Arch. Psychol., No. 102.

Wilson, G. M., 1950. Arithmetic, in *Encyclopedia of educational research,* ed. W. S. Monroe, New York, The Macmillan Co.

Wolfle, D., 1940. *Factor analysis to 1940,* Psychometric Monogr., No. 3, Chicago, University of Chicago Press.

Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt & Co.

Woodworth, R. S., 1947. Reinforcement of perception, *Am. J. Psychol., 60*:119–124.

Woodworth, R. S., and Sells, S. B., 1935. An atmosphere effect in formal syllogistic reasoning, *J. Exper. Psychol., 18*:451–460.

Chapter 9

Judgment: Stimulus Variables

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

—Robert Frost

WHEN the problem to be solved is a complex one that leads the thinker through a variety of intellectual activities, the last phase of the problem-solving episode is usually a judgmental or decisive process. The thinker halts his productive activity to judge the merit of what he has produced. When the problem is a simple one that consists merely of a choice between two alternatives, the judgment is all there is to it: Shall I take apple pie or cherry pie? Any account of intellectual activities that pretends to completeness must take up these activities by which people settle uncertainties, draw conclusions, form decisions, make judgments. The term judgment, although it has had many meanings throughout its long and honorable career, serves very well at the present moment in the history of psychology as a name for this conclusive activity.

The judgment process, like most others, was first studied subjectively, and some information of interest may still be obtained by this method. More dependable quantitative information comes from psychophysical experimentation, especially in recent times from the method called absolute judgment, or single stimuli. Experiments on

affective or esthetic judgments have contributed also to the psychology of judgment in general, particularly to the influence of scales of judgment.

As psychologists became interested in ratings of personality characteristics, they found it necessary to borrow some psychophysical methods and to invent some new ones. Incidental information about the judgment process, e.g., the "halo effect," may be gleaned from research on judgments of people. The psychology of judgment has acquired additional breadth from methodological studies of attitudes and opinions, and the arena of public opinion offers an excellent proving ground for hypotheses derived from laboratory experiments.

When psychologists report the results of their experiments on these and other topics, the term judgment is often used in a noncommittal way. The experiment may be directed toward perceptual or social processes, but judgments are what are tabulated and used in the analysis of the results. The term is not explicitly defined, and the judgment process as such is usually ignored; there is as yet no well-recognized psychology of judgment. It is the purpose of the present chapter and the next to pull together the evidence from scattered sources, showing how the judgment process meshes with other thought processes, and where it fits into psychology in general. Some details not included in the present account are available in an earlier attempt at integration (Johnson, *1945*). We shall see, when the terminology is clarified and the evidence systematized, that the psychology of judgment has a respectable body of facts and principles of its own. In fact, since the judgment process is relatively simple, as compared, e.g., to creative thought, this topic is really in a rather advanced state of scientific maturity. The evidence is considerable, but a convenient division can be made by considering some introductory points and stimulus variables in this chapter and response variables in the next chapter.

THE FUNCTION OF JUDGMENT

It is unfortunate that psychological experiments end so quickly. The subject makes 200 judgments of lights or designs or attitude items, then he goes home, apparently unaffected by the experience. In fact, he is often lured into the laboratory with the assurance that nothing that happens therein will have any effect on him. Judgments made at the choice-points of life outside the laboratory, where they

are often called decisions, may be similarly inconsequential, but some have permanent consequences. When anyone decides upon one course of action rather than another—decides to get married, to quit a job, to commit murder, to buy a hat, to borrow money, or to take the road less traveled by—the remainder of his life may be thereby altered. Some decisions, once executed, are irreversible, like a decision to have a tooth pulled, while others merely outline a course of action and initiate a learning process, like a decision to give up smoking. Novelists, playwrights, clergymen, lawyers, and fortune tellers have concerned themselves more than psychologists with decisions of great consequence—with a few recent exceptions. The degree of similarity between judgments made in the psychological laboratory and those made in the hurly-burly outside the laboratory walls is unknown, but certainly any general principles of judgment that can be derived from laboratory experimentation will have some degree of applicability to the events that occur at the crossroads of life's decisions.

In general judgment is a conclusive or decisive process, not a productive one, that brings a thoughtful episode to an end. In the case of highly differentiated human beings it is a retroflex process in that the thinker takes into account the motivational and instructional conditions that initiated the thoughtful episode. The preparation sets up two or more alternatives, between which a choice is made. These alternatives may be perceived directly in the environment, or they may be recalled from the past, or they may be produced *de novo* by a creative process. While many decisions are inconsequential, some, when executed, decide the thinker's destiny.

Logicians, discussing judgment, are interested in the relation between the outcome of thought, as announced in the form of a conclusion, and the data upon which the judgment is based. Judgment has the function of evaluating or settling an uncertain state of affairs. They use such terms as premise, proposition, data, and inference. Psychologists who emphasize perceptual organization are likely to say that judgment has the function of clarifying an unstructured situation; they use such terms as organization, clarity, and ambiguity. After judgment things look different. Others, emphasizing behavior, might speak only of response, patterns of response, and courses of action. We shall find all these terms helpful in various contexts, and need not commit ourselves to a preference at this point. As a neutral phrasing we may speak of the raw material or stimulus material, a result of

perceptual, recall, conceptual, affective, or other productive processes, upon which the judgment process operates.

CLARIFICATION OF TERMS

Because confusion in terminology is so common, and because most of the concepts under consideration are quite abstract, it is worth while to clarify the meaning of judgment by comparing and contrasting judgment with several related functions. Nothing new is intended here; but present usage should be made explicit.

JUDGMENT AS OVERT COMMUNICATIVE RESPONSE. Inspection of published reports of psychological experiments will show that when the term judgment appears, it usually refers to a response that is a consequence of, or overt expression of, or a way of communicating the results of some other process. Perception and feelings, for example, are not directly observed. They may be inferred from casual behavior or speech or, for greater precision, from judgments made under controlled conditions. In psychophysical experiments the response observed by the experimenter is a word, *heavier* or *lighter,* or a mark in one box or the other. The response is in the form of a judgment, but the experimenter is not interested in the judgment *per se.* Rather he records the judgment as communicating the underlying perceptual process. The rationale of the psychophysical methods is in large part an attempt to control the peculiarities of the judging process, or to balance them out statistically or experimentally, so that the overt communicative responses can be clearly related to the perceptual processes. An important part of psychophysical methodology is the establishing of an optimal preparatory set that standardizes the relation between stimulus and response for all subjects under all experimental conditions. The judgment process cannot be entirely standardized, however, hence some of our best evidence on judgment comes as a by-product from psychophysical experiments and from experiments on affective tone and attitudes. Furthermore, these experiments can be redesigned so that the judgment process, instead of being minimized, is emphasized and brought more prominently into view.

JUDGMENT AND PROBLEM SOLVING. Judgment tasks are those in which two or more responses or courses of action are specified, the judge being required to choose between them. "Which is more delightful, the smell of a rose or its appearance?" "Which line is longer?" We are likely to use the term problem solving for the

activities involved in producing a solution that is specified only partially, production being a large part of the problem. "Plan a garden for a small backyard." "Take a man's vest off him without removing his coat." If we use thought as a generic term for the solution of problems, then judgment is a kind of thought and *a* judgment is a kind of problem the solution to which is a response in one of several categories prepared in advance of the judgment. Because of the wheels-within-wheels character of problem solving it often happens that alternative solutions are produced during problem solving, then the thinker judges the worth of each. Or the thinker produces various preparations and judges the utility of each in relation to his overall motivation before committing himself to any one.

JUDGMENT AND ESTIMATION. Judgments of the heights of trees may be expressed in two categories: *tall, short.* They may also be expressed in several categories designating adjacent regions of a scale: *very tall, tall, medium, short, very short.* When the judgments are expressed on a continuous scale, e.g., one of inches and fractions of an inch, the term estimating may be used, as a special case of judging. Because the scale used in estimation is continuous, there are no boundaries between categories and the phenomena of judgment that occur near category boundaries do not occur. Otherwise estimation may be subsumed under judgment.

JUDGMENT AND PERCEPTION. Judgment is, of course, not the same as perception. If one perceives the situation clearly enough that the activity under way may proceed, no distinguishable act of judgment occurs. But if the perceptual field is not clear, if the first perception is not rewarded, choosing one of the alternatives may be called an act of judgment. So stated, the distinction seems definite, but if we examine it more closely, the distinction becomes less sharp. In Figure 9.1

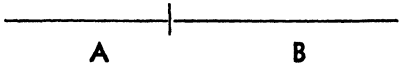


FIGURE 9.1.

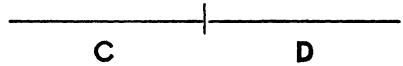


FIGURE 9.2.

which segment of the line is longer, A or B? B is immediately perceived as longer, without deliberation. In Figure 9.2 which is longer, C or D? In this case a difference is not clearly perceived. The observer hesitates, puts some thought to the task. Then he may say: "I am not certain, but I would judge C longer than D." Consideration of these

two examples and the infinite number of intermediate examples suggests that the conditions that give rise to judgment are not sharply differentiated from those that permit immediate perception. The difference, as far as we know, is a quantitative one.

Three characteristics of the response permit a quantitative comparison of judgment and immediate perception. It has been known for years that when people are asked to compare two lines, using one as a standard and calling the other *longer* or *shorter,* an increase in the physical difference between the lines will produce (1) an increase in the percentage of judgments in one category, (2) an increase in the confidence reported for the judgments, and (3) a decrease in the time required for judgment (Fullerton and Cattell, *1892;* Johnson, *1939;* Cartwright, *1941;* Festinger, *1943*). The details are given in the next chapter. Presumably as confidence approaches a maximum and time of response approaches the reaction time, judgment is minimized and the act approaches direct perception. Hollingworth's (*1925*) differentiation of perception and judgment depends on just this quantitative difference.

JUDGMENT AND RECALL AND RECOGNITION. The difference between recall and recognition neatly illustrates the difference between production and judgment. In the former responses are produced to fit specified conditions; in the latter a choice is made between alternatives at hand. There is some empirical justification for asserting that recognition is a form of judgment because individual differences in confidence and speed of recognitive judgments correlate highly with individual differences in confidence and speed of other kinds of judgments (Johnson, *1939*).

PREPARATION FOR JUDGMENT

Since judgment is a straightforward process, with definite antecedents and consequences, it lends itself to experimental control, and analysis into independent and dependent variables is profitable. The principal independent variables are the preparation for judgment and the stimulus materials upon which the judgment is made.

The preparation for judgment, like the preparation for any other activity, may be established in several ways. In the laboratory the set is usually prepared by verbal instructions. "Which of these two lines is longer?" "You are to judge the attractiveness of these colors without regard to their use." "Rate these men in respect to their sociability."

The set may be reinstated from previous experience, as in the case of the inspector who comes to the factory in the morning and judges articles coming off the production line just as he did the day before.

The set may also be a task-induced set, in part at least. When a young child is asked to arrange several blocks in order of weight, the instructions may not establish a set at once because the child has not previously concentrated solely on that dimension. As he gets into the task, hefts the blocks, and notes variations in weight, it becomes easier to get set for this dimension and ignore others. But if the task is very complicated, establishment of the set may be a productive task in itself, antecedent to judgment proper. The set for judgment may also be established by a biological drive, as in the cafeteria situation when one has to decide what to spend his money for. Anyone who is really hungry has a different set, is sensitive to different sensory stimuli, than one who is eating because of habit or social motivation.

The preparation for judgment, however it may be established, accomplishes three things. The thinker is alerted, prepared for action at the appropriate time. The stimulus objects, data, or materials of judgment are specified; i.e., the thinker is sensitized to, or set for, some aspects of the environment or of memory rather than others. Third, the form of the response or the alternative response categories are specified. This judging function is carried out by living organisms, of course, and the little preparatory sets operate within the organism's overall motivation. And naturally the effectiveness of any set in coördinating the organism's activities varies from moment to moment and from person to person.

Although the response expressing the judgment is, by definition of judgment, specified with some precision by the preparation, precision of specification of the things to be judged varies considerably. Thus it becomes necessary to differentiate, in respect to complexity of the stimulus material, two kinds of judgment: simple and complex. By simple judgments we mean those in which the subject judges a simple, clearly specified aspect of the stimulus situation, such as the length of a line or the pitch of a tone. Complex judgments are those in which the object of judgment is complex and the stimulus aspect or aspects to be judged are not distinctive. When a foreman is asked, for example, to rate the merits of his crew or when a conscientious voter tries to select the best candidate for public office, the stimulus aspects to be judged are not at all simple and well specified. The response variables

for these two types of judgment can be treated together, in the next chapter, but the stimulus variables must be treated separately.

JUDGMENTS OF SIMPLE MATERIAL

Assuming that the judge is now prepared to judge, the important question becomes: What does he judge? The other large class of independent variables of judgment, in addition to the preparation, is the raw material, the stuff of which judgment fabricates a decision. The two variables are not independent of each other, however, for the preparatory set determines what the judge attends to and, in reverse, variations in the stimulus material facilitate the adoption and maintenance of some sets rather than others.

If the judgment is a simple one, like judging which of two fish is larger, or whether the toast has browned to the desired shade, the preparation emphasizes just one aspect or dimension of the stimulus objects, e.g., the size or the color. In many judgments of the laboratory, shop, kitchen, and playground the objects being compared or rated vary principally along a single dimension or continuum, and it is this aspect of the objects that is judged. The instructions and the simplicity of the arrangement of the stimulus objects make it relatively easy to isolate and attend to the pertinent aspect or quality of the objects of judgment. Experiments on these simple judgments have been directed chiefly toward three characteristics of the stimulus objects: perceptual, affective, and abstract.

JUDGMENTS OF PERCEPTUAL CHARACTERISTICS

Some of the most valuable and precise information on our topic comes from judgments of directly perceived aspects of the stimulus objects. In these the observer attends to and judges some specified attribute of the stimulus objects, e.g., color, loudness, distance, or size. Methods of experimentation and treatment of the results, called psychophysical methods, e.g., average error, constant stimuli, etc., have been well worked out, and this field of psychology is a mature science. The high degree of precision characteristic of psychophysics is due to good preparation of the subject and good control of the stimulus variables. Sounds, lights, weights, and other stimuli can be accurately measured in standard physical units, and can be presented repeatedly for judgment. In all sensory departments a wide range of

stimuli is available, to be spaced along the stimulus continuum at the experimenter's convenience.

JUDGMENTS OF AFFECTIVE CHARACTERISTICS

Experiments on affective judgments have contributed a large share of present knowledge of the psychology of judgment. In these the observer is instructed to report on his likes and dislikes, how pleasant or unpleasant something looks or smells to him, how intensely he wants something, how strongly he reacts for or against something. Common materials for affective judgments are colors, musical intervals, faces, slogans, symbols, familiar activities, and foods.

Most people enjoy directing their attention toward the affective characteristics of any object, and casual conversation runs often to expressions of likes and dislikes. This may be the most primitive, effortless form of reaction to the environment. At any rate a set for these personal judgments is very easily established, and may operate when the instructions call for more impersonal or objective judgments.

Some of these hedonic stimulus objects can be well controlled and suitably varied. Colors, musical intervals, and the chemical composition of certain foods can be accurately measured in standard units, conveniently spaced along a stimulus continuum, and presented repeatedly for judgment. Stimulus control is more difficult for other objects, e.g., faces. There is no standard unit for the attractiveness of faces, though it is possible to arrange faces on a socially derived continuum of attractiveness. Methods are available, e.g., order of merit, rating, pair comparisons, and the like, and this field of psychology is fairly well advanced.

JUDGMENTS OF ABSTRACT CHARACTERISTICS

A third class of judgments may be distinguished in which attention is directed to abstract or conceptual properties of the objects of judgment, and the judgments are made on this basis. In a library, for example, the cataloguer who has to decide whether to list a new title, *The Russian People,* under anthropology, current events, political science, or subversive influences is making such an abstract judgment. His judgment is directed toward abstract characteristics of the book and their relations to an abstract classification system, not toward color, size, or attractiveness. The judge in a court of law who rules on the admissibility of evidence supplies another illustration. In studies of

personality judges are often asked to rate people on such abstract qualities as social usefulness and honesty and to rate occupations on a scale of prestige. Judging the conclusion of a syllogism is another example. Most questions of grammatical usage and most questions on tests of educational achievement are in the form of abstract judgments. The fact that these are abstract judgments has not been emphasized because the judgment process is presumably held constant, the important independent variable being knowledge of the things being judged. Therefore abstract judgments merit more discussion at this point than the other two varieties.

Although no clear-cut comparisons have been made, we may infer from the reluctance of many people to undertake abstract judgments, and the errors and confusions that commonly occur, that abstract judgments are more difficult than the more primitive perceptual and affective judgments. The evidence on concept formation summarized in the previous chapter is pertinent. Intellectual operations directed toward abstract characteristics are more difficult simply because the thinker has more to do. Abstracting a trait such as sociability from a jumble of past and present impressions that occurred in heterogeneous contexts, and differentiating this trait from cheerfulness, talkativeness, intelligence, and aggressiveness, is more involved, requires more thought control, than isolating length from width and color, or feeling tone from objective dimensions. There are more possibilities for interference because things are similar and dissimilar in respect to so many abstract properties, one of which may be about as prominent as another. It seems likely, therefore, that most people have to work harder at abstract judgments, with less rewarding consequences, other things being equal. The set that controls thought is organized with more effort, and more effort is required for maintenance and resistance to interference.

Refined quantitative research on abstract judgment is meager and some may object to the classification of these activities as judgments along with perceptual and affective judgments. But the function of the judging process and its relation to other intellectual processes seem to be the same as in the other two varieties. The raw material of judgment, instead of resulting from perceptual or affective processes, is conceptual, but two concepts may be studied and judged *same* or *different* just as two sugar solutions may be tasted and judged *same* or *different*. As we shall see in the next chapter, individual dif-

ferences in speed of perceptual judgments correlate quite closely with individual differences in speed of abstract judgments. There is some generality of individual differences in confidence in judgments of various types of material also.

THE PROBLEM OF STIMULUS CONTROL. If we wish to put abstract judgments into the same experimental design as the others, the first problem is one of stimulus control. Our understanding of perceptual mechanisms has been facilitated by the availability of a wide range of stimuli, accurately controlled and suitably spaced along the stimulus continuum. It is much harder to arrange objects of judgment at known locations on a stimulus scale, and memory usually prohibits repeated presentation. It is true that vocabulary tests of the multiple-choice type deal with abstract judgments, but the items are standardized merely in respect of difficulty of judgment. The dimension of similarity that makes the judgment easy or hard is usually not known.

The development of scaling procedures for a Thurstone-type attitude scale (see Guilford, *1950*) has two important implications for the psychology of judgment of abstract attributes. First, the success of these scaling procedures shows that the judges who aid in the scaling of the items—educated adults with an interest in the task, not the subjects who take the test—are able to conceive an abstract stimulus continuum, e.g., favorableness of the item toward communism, to compare items in this respect, and even to sort the items directly in accordance with their location on the abstract continuum. Many judges are able to maintain a set toward this abstract aspect of the item in spite of interference from other aspects, foreign to the attitude continuum, and even from the personal acceptability of the item (Ferguson, *1935;* Hinckley, *1932;* Pintner and Forlano, *1937*). But highly involved judges, e.g., Negroes scaling items about Negroes, do not give the same scale values as disinterested judges (Hovland and Sherif, *1952*). The scale values derived from abstract judgments made in various ways are rather closely related (Ferguson, *1940;* Saffir, *1937;* Farnsworth, *1945*), a fact which rules out the possibility that these scale values are artifacts.

Second, these scaling procedures solve the problem of stimulus control for abstract judgments. They make possible the manufacture of stimulus material to definite specifications in regard to ambiguity and

location on a conceptual continuum. While these scaling procedures have had their chief application in the construction of attitude scales, there is no good reason, a priori, why symbols representing various degrees of any abstract quality at all, however subtle, cannot be scaled in the same way. A minor limitation is that the scale values are derived from the consensus of a group and may not be precisely valid for any one individual.

Abstract material prepared in this way has been used in a few experiments on judgment. Chant and Salter (*1937*) used Droba's scale of attitude toward war to show the relation between difficulty of a judgment and the galvanic skin reflex. Johnson (*1940*) used such material to study the relation between confidence in a judgment and distance from the category threshold. Cartwright (*1941*) had his subjects scale names in respect to radicalism-conservatism in order to get stimulus material for a study of decision time. McGarvey (*1943*) had her subjects scale occupations in respect to prestige and certain forms of behavior in respect to undesirability in order to study the anchoring of the scales. Mosier (*1941*) has scaled over 200 adjectives on a scale of favorableness-unfavorableness. All of these studies have contributed evidence which will be presented later.

Judgments of other stimulus characteristics could be described, but these three are the main varieties in which the subject is prepared to respond to variations along a specified continuum and to ignore other variations. The three varieties of judgment are similar in that one process, perceptual, affective, or conceptual, emphasized by the preparatory set, delivers the raw material for judgment. The judgment process may be held constant and the other processes emphasized if information on perceptual, affective, or conceptual processes is desired. Or the judgment process may be experimentally emphasized by suitable control of the materials of judgment.

Judgments of these simple kinds, with known stimulus correlates, permit the discovery of precise relations between the dependent variables of judgment, e.g., category of response, time, and confidence, and the independent variables, e.g., the preparation and the stimulus materials. Research of this kind has led to an understanding of such general principles as central tendency and anchoring effects. The general principles of judgment, to be described below, apply, of course, to judgments of all kinds of material.

JUDGMENTS OF COMPLEX MATERIAL

Most of the judgments of real life are not simple. By a complex judgment we mean one in which the stimulus material is heterogeneous, with no one prominent aspect or dimension to which the judge can be easily prepared to respond. The response is therefore correlated with more than one aspect of the stimulus material.

Complex judgments are made in many practical situations, as when a cook decides that a cake should or should not be removed from the oven, when stockmen judge cattle, when teachers grade their students' papers, when anyone decides to buy a car, and when a jury judges a defendant guilty or innocent. The teacher's preparation for grading students' papers, for example, may be quite general and indefinite, an unanalyzed result of years of experience. It may, however, be quite explicit. He may write down the factors he is looking for: content, organization, originality, grammar, punctuality, etc. He may even assign a weight to each. But usually he will not judge each paper on the basis of each criterion one at a time. The judgment will usually be an overall or general judgment, based on all these criteria, and perhaps some others not clearly specified, in unknown amounts. It is an empirical problem to identify the stimulus variables which determine the judgment and assay the relative weights of each.

If the judgment occurs at the end of a creative problem, the preparation is still less specific. Someone writes a poem, then rereads it for the last time, trying to decide whether to send it to a publisher or tear it up. The preparation for judgments of this sort is self-imposed. The poet tells himself what to look for, what to ignore. He reminds himself of his standards and his desire for publication. He may attempt to put himself in the position of an editor or critic and adopt another's standards and point of view. He may have difficulty shifting from a productive or creative set to a critical set, but sets become habits in time and even the shifting of sets may become easier with practice.

A general description of the preparation for complex judgments would have to begin with the motivation for the whole enterprise of which the judgment is a part. Whether the motivation is personal or social, primary or derived, a general overall preparation to make certain responses, e.g., *yes* or *no, high* or *low,* is established. But if the stimulus material is complex, and the general set does not further progress toward the goal response, the judge surveys the situation and,

deliberately or otherwise, prepares a more specific set to respond to certain aspects of the situation, certain attributes, dimensions, or traits. He may deliberately take time to set up the criteria of judgment, isolating those attributes that he considers important, or he may hurry along, not knowing the basis of his judgments. He may refuse to analyze, preferring a global impression. In any event his criteria of judgment are related in some way to progress toward his goal. This is what gives complex judgment its retroflex character. The material of judgment is produced under the guidance of a specific set, but is later evaluated in reference to the general overall set which started all this.

Often the criteria to be used in making the judgment are specified by instructions. The personnel manager may instruct the foremen on what variables his men are to be rated. The jury is instructed what evidence is admissible and what is not.

Whether a judge follows the instructions that he gives himself or that others give him is an empirical question. A priori, one would not expect that any little detail of the judging process would override his overall motivation. Nor would one expect that anyone's judgments will be influenced by criteria which he cannot disentangle from an inclusive configuration. Considerable research, in several fields of psychology, has been directed toward identifying the variables of the stimulus material that affect the judgment and estimating their relative weights.

IDENTIFICATION AND WEIGHTING

When we tried, in Chapter 6, to uncover the variables determining the response in free association, we found (1) that in addition to the stimulus word the preparatory set is important, and (2) that the preparatory set itself can be studied by tabulating the response words and analyzing the relations between each response word and stimulus word. Now the complex-judgment experiment is a mirror image of the free association experiment. In free association the preparation determines the aspect of the situation to which the subject attends, namely, the stimulus word. But in the complex-judgment experiment the object of judgment is heterogeneous, and just what the subject attends to is not precisely determined in advance. In free association the response is not determined, to any extent, by the preparation but may vary freely. But in the judgment experiment the response is

closely restricted to one of a small number of categories specified in the preparation. To put it briefly: in the one experiment the overall obvious preparation fixes the stimulus and permits the response to vary; in the other the effective stimulus varies and the response is fixed. In both cases if we wish to know more about the details of the preparation we must analyze the relations between response variables and stimulus variables in more detail. If we wish to identify just which aspects of the complex stimulus the judge takes account of and which he ignores, we must tabulate the responses and compare them with various likely aspects of the stimulus. This, of course, is the standard method of experimental psychology, of experimental science in general. If a complex object of judgment, e.g., a human being, is to be judged in one of several specified categories of response, and this object can be judged in many respects, e.g., height, sociability, age, income, talkativeness, etc., a comparison of variations in these respects with the responses will disclose what respects the judge is attending to when he makes his judgments. From research of this kind a few general principles of judgment can be extracted.

1. Our first principle, though perhaps obvious, is of strategic methodological importance. *The judgment may be influenced by stimulus aspects or variables to which attention is not directed by the explicit instructions or by logical implications of the stimulus material.* The irrationality of some men has been pointed out by their presumably more rational critics for as long as man has been called a rational animal. The principle may be illustrated by an experiment in judging syllogisms (see page 245), which also makes a neat contribution to methodology.

Woodworth and Sells (*1935*) guessed that judgment of the conclusion of a syllogism depends on the impression or atmosphere created by the premises as well as by the logical relations of the premises. Negative premises set up a negative atmosphere. Affirmative premises set up an affirmative atmosphere. And conclusions in agreement with the atmosphere thus engendered are likely to be accepted. There are complications, but specific predictions can be made from the atmosphere of the premises. Sells (*1935*) presented syllogisms with a large percentage of invalid conclusions to a group of educated adults and found that of these invalid conclusions those in agreement with his predictions were accepted much more often than the others.

In these experiments the response side of the subjects' set was ade-

quate. But they were attending to the wrong aspects of the stimulus material, to the atmosphere of the premises rather than to logical relationships. One might say that, although they were firing right, they were aiming wrong. And the experimenters' technique enabled them to discover what the subjects were aiming at.

Many subjects, when judging the items of printed psychological tests, attend to the position of the alternatives on the page, and the like, as well as the abstract significance of the item (Cronbach *1946, 1950*). When the responses are properly tabulated, it appears that some subjects prefer *True* and *Yes,* while others prefer *?* and *Indifferent.* Such sets can be modified by instructions which encourage the subjects to redefine the stimulus material.

2. Our second principle, somewhat similar, is that *the judgment may be influenced by stimulus variables which the judge cannot or does not report.* Under some conditions people can respond with accuracy above chance levels to subliminal stimuli, i.e., to stimuli which have been reduced in intensity below the threshold of report (Baker, *1937;* Miller, *1939;* Williams, *1938*). It is quite likely that our immediate intuitive judgments of personality are based in part on such subliminal cues, subliminal in a perceptual sense or in the sense that they depend on memory impressions too faint to be reported. Repression may augment this effect, particularly in judgments of personality. Mr. A may not recall a distasteful experience with Mr. B even though his judgment of Mr. B is influenced by that experience.

3. These first two principles are stated negatively. The next two, referring to social influences on judgment, can be stated positively. *Most judges give extra weight to stimulus material, or suggestions, attributed to people whom they regard as experts* (Asch, Block, and Hertzman, *1938;* Lorge and Curtis, *1936;* Saadi and Farnsworth, *1934*). This principle can be reversed. The effect of the expert's suggestions in shifting judgments can be used as a measure of the prestige of the expert (Lurie, *1938*). It is even possible to manufacture a temporary expert in the laboratory by making one judge appear more successful than the others (Kelman, *1950*). When controversial material is communicated to an audience, it seems to be more effective if the communicator draws the conclusions than if the audience draws its own conclusions (Hovland and Mandell, *1952*).

4. Analogous to the influence of the expert is the influence of the majority. *Most judges give extra weight to stimulus material, or sug-*

gestions, attributed to the majority of a group with whom they identify (see Sherif, *1936*). Obviously, expert and majority opinions may conflict, and attempts have been made to determine which exerts the more powerful influence on judgments of controversial issues under different conditions (Burtt and Falkenburg, *1941;* Marple, *1933;* Moore, *1921;* Berenda, *1950;* Asch, *1951*).

5. Our fifth generalization is not so well established as the first four, but it summarizes the results of several experiments. *When the judgment called for is avoided, because of its difficulty or for any other reason, judgment of some other stimulus aspect will be made.* When the judge is prepared, usually by instructions, to judge a complex object in respect to one stimulus attribute and such a judgment is difficult, he could simply withhold judgment. And some people do. But most people go ahead and make a judgment anyway. Comparison of the responses with the possible stimulus attributes often shows that they have shifted to judgment of some other stimulus attribute more prominent than the one they were prepared to judge. For example, Thornton (*1943*) demonstrated that, when college students are shown a photograph and asked to rate the person as to various personality traits—an almost impossible task—a person smiling is likely to be rated higher in honesty than the same person not smiling, and a person wearing glasses is likely to be rated higher in intelligence, dependability, industriousness, and honesty than the same person not wearing glasses.

6. The sixth principle is a special case of the fifth. *When an abstract judgment is called for, an affective judgment is commonly given.* Maintaining a set toward abstract properties of a complex object of judgment in spite of distractions is difficult and tedious for many people, while affective judgments, such as "I like it" or "I don't like it," are so much fun. Thus when anyone is asked whether a strong labor union increases economic stability, he is likely to give a personal affective response to the symbol "labor union" instead of judging the abstract question as asked. This phenomenon is regularly observed in judgment of the controversial issues of the day, because they are abstract and difficult, and the content is usually of affective significance. Evidence from laboratory judgments and from opinion polls is also available (Cantril, *1938;* Lund, *1925;* McGregor, *1938*).

Janis and Frick (*1943*) made use of syllogisms to test the influence of affective involvement on abstract judgment. Their instructions to

the subjects—which are particularly important in experiments of this kind—and the first two of their syllogisms follow.

Instructions: This is a test in reasoning. You are to check each of the following arguments as "sound" or "unsound." A sound argument is one in which the conclusion follows logically from the premises. (Note: "All" in logic means *each and every* case. "Some" means *at least one* and perhaps *all* cases.)

1. No Bolsheviks are idealists and all Bolsheviks are Russian. Therefore, some Russians are not idealists.
2. The Eskimos are the only people who eat nothing but meat and it is found that all Eskimos have good teeth. So we may conclude that no people who eat only meat have bad teeth.

After judging 16 of these syllogisms each subject reread the conclusions and was asked whether he agreed or disagreed with each one. Thus the judgments could be sorted into four types:

agree-valid
disagree-invalid
agree-invalid
disagree-valid

The effect of the attitude would be expected to cause the subjects to judge the conclusion valid when they agreed with it and invalid when they disagreed with it, that is, to increase the proportion of judgments in the first two of the above categories and decrease the proportion of judgments in the last two.

Of course if the subjects judged all syllogisms correctly, no effect of attitude toward the conclusions could be detected. But syllogisms are difficult. Even though the subjects, who were graduate students with no formal training in logical analysis, had only two alternatives to choose from, they were wrong on 23 percent of the items. And, as expected, there was a preponderance of errors of the agree-valid and disagree-invalid types. Although only 19 subjects took part in the experiment, and only 70 errors were analyzed, the results are statistically significant. With subjects of less education and interest in abstract things, the difference would probably be even greater.

Lefford's (*1946*) experiment is similar in some respects. He asked several groups of college students to judge the validity of syllogisms, instructing them "to judge whether the conclusions arrived at in these passages are warranted by the statements given in support of them."

Half the syllogisms were concerned with controversial issues, as in the two examples below.

15a. War times are prosperous times, and prosperity is highly desirable, therefore, wars are much to be desired.
16a. All communists have radical ideas. All CIO leaders have radical ideas, therefore, all CIO leaders are agents for communism.

The instructions call for abstract judgments of the validity of the conclusions as based on the given premises. Such judgments are difficult, even for college students, and the scores were quite low.

The other half of the syllogisms were matched for form, but the content was noncontroversial, not affective.

15b. Philosophers are all human, and all human beings are fallible; therefore, philosophers are fallible too.
16b. All whales live in water; and all fish live in water too; therefore all fish must be whales.

These syllogisms, of the same form as the others and approximately the same length, were judged correctly much more often by college students. Apparently the difficulty of the judgments of the controversial syllogisms is due to the intrusion of affective material, or, to put it more precisely, the subjects were set for abstract judgments but could not avoid making affective judgments.

The effect of order of presentation on the number of errors is particularly enlightening. The group of 59 students who judged the controversial items first did poorly on both kinds, while the 30 subjects who did the noncontroversial items first did considerably better on both. If the subjects in these two groups were of the same ability, the most plausible interpretation is that practice on the noncontroversial items improved subsequent performance on the controversial items and practice on the controversial items interfered with subsequent performance on the noncontroversial items.

7. The fifth principle, and therefore the sixth also, can be reversed. *When the judgment called for is straightforward and easy, the judge is likely to maintain his set for that judgment and not be diverted into judgment of something else.* Evidence for this statement comes from experiments indicating a limitation on the effects of suggestion (Coffin, *1941;* Lewis, *1941;* Birch, *1945;* Berenda, *1950*) and also on the effects of the judge's desires (Cantril, *1938;* McGregor, *1938;* Cronbach and Davis, *1944*) when the situation is well structured, i.e.,

when the called-for judgment is easy. "Response sets" show up more clearly on psychological tests when the most successful subjects are eliminated (Cronbach, *1950*). It was probably for this reason also that Sells (*1936*) eliminated subjects with training in logic from his experiment on the atmosphere effect. In a study by Thornton (*1944*), following up the one mentioned above, he found that the effect of wearing glasses on judgments of personality was smaller when the actual people were seen than when photographs of these people were judged. As he puts it, these factors "will have decreasing effect upon judgments of personality traits as the number of other cues upon which judgment may be made increases."

This seventh principle, which is probably the most important of all for practical purposes, is illustrated also by the "good judgment" produced by training. Trained social scientists do not make the errors in judging controversial social issues that are made by intelligent untrained citizens. We have noted above that attitude items can be scaled quite well by judges of diverse sympathies when properly instructed. When teachers and school children rate behavior items, average scale values for the two groups are very similar in spite of opposing affective involvements (Witryol, *1950*). Precision of the individual children's judgments improves with age. Ratings of personnel by foremen and supervisors are improved by short training courses. And ratings of behavior by research psychologists sometimes reach high levels of accuracy after serious practice. In general, improvement in interjudge agreement is obtained by getting judges to agree on the definition of the variables to be judged, and training them to attend to these and to avoid the distraction of other, more obvious variables.

Training in judging anything consists largely of practice in identifying the important variables in complex objects of judgment, e.g., cattle. Good and bad specimens are examined. The expert points out what is important and what can be ignored. Weights are assigned to these variables. Official committees often develop score cards that make the task easier. There is much to be learned by the student of the psychology of judgment in studying a textbook on judging dairy cattle (e.g., Harrison, *1940*) or observing stock judging contests.

In one study of the reliability of judgment (King, Ehrmann, and Johnson, *1952*) two observers rated nursery-school children, one at a time, for about an hour each, on such variables as sociability, aggres-

sion, and forwardness. Interjudge reliability was low. Then the two judges worked together, defining their variables and differentiating them from extraneous variables. After several practice sessions of this kind independent observations reached a high level of agreement (see Figure 9.3).

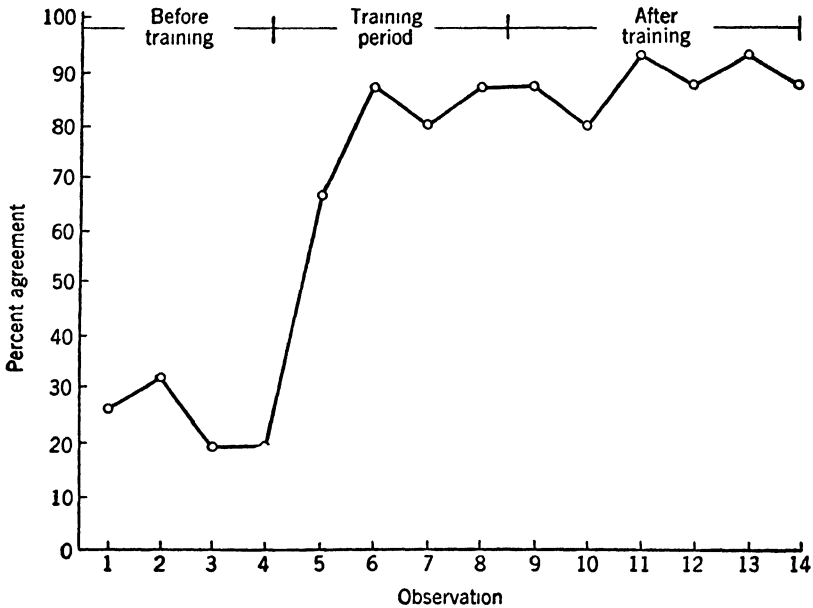


FIGURE 9.3.The effects of training on accuracy of judgment. The ordinate shows agreement between two observers rating nursery-school children on several variables of social interaction. From King, Ehrmann, and Johnson (*1952*).

ESTIMATING RELATIVE WEIGHTS. These general principles merely identify factors that influence the judge as he makes his decision. Some attempts have been made to go beyond these generalizations and estimate how much weight each factor exerts on the judge as he ponders the problem. Under some conditions people will be influenced more by majority opinion than by expert opinion, and under other conditions the relative weights of the two are reversed (Burtt and Falkenburg, *1941;* Marple, *1933;* Moore, *1921;* Berenda, *1950*). In one real-life study of voters' judgments of candidates for the office

of state senator (Hartmann, *1936*), emotional campaign material was more influential than rational material, but we know from the evidence assembled under Principle 7 that this is not always true. The stimulus material and the special conditions of the judgment seem to be more important than any general set or attitude toward one type of stimulus material rather than another. And we know that training in dealing with any particular kind of material reduces the influence of suggestion and emotional factors on judgments of that kind of material.

The methodology of this kind of research is important. It is comparatively easy by several methods to determine that such and such a factor has some bearing on the judgment. To determine in any general sense the relative weights of two or more factors is much harder. For example, to compare judgments made after exposure to emotional and to rational material, or judgments based on knowledge and on desire, it is necessary to assume that each kind of material used in the experiment is an equally potent representative of its class; i.e., that the rational material is as rational as the emotional material is emotional. If this assumption is not justified, the conclusions are restricted in generality. It is now possible by the scaling procedures mentioned above to make a more precise approach to these problems. If the stimulus material can be located on a scale of rationality or desirability or authoritativeness, a wide range of stimulus material can be prepared, and refined comparisons or correlations can be made.

The public opinion poll offers possibilities for indirect assay of the weight of various factors in determining judgments of social issues. Objective data, such as educational and economic status, are rather crude, psychologically; but, if the sampling is properly done, the full range of variation of these factors is included in the relative weightings. Thus the results are more important than those obtained from a restricted sample like college students. Williams and Mosteller (in Cantril, *1947*) phrase the question thus: "When social decisions are made by individuals who have disparate income and educational backgrounds, which of these factors plays a greater part in the molding of their judgments?" This particular question is a difficult one because education and economic status are correlated. But by comparing the judgments of various political and economic issues with educational status of the respondents and again with their economic status, Williams and Mosteller were able to uncover some general

trends. "Education is found to be more important where factual knowledge gives an insight into implications, for example, in determining opinion toward a negotiated peace or toward post-war problems. . . . Economic status, on the other hand, determines opinion more than education on all issues where the chief impact or where the clear-cut implications deal with financial return."

This study does not carry us far but it does show something of the socioeconomic cross currents that sensitize people at large to this or that aspect of the contemporary scene. More analytical research of this kind and two-way breakdowns of opinion data (Cantril, *1947,* Chap. 13) would contribute to general principles of judgment as well as to practical information about the weight of specific factors. In fact the opinion poll offers a good proving ground for hypotheses about thought derived in the laboratory or clinic.

When data on judgment suitable for correlational analysis can be obtained, the correlation approach offers a sharper method for evaluating the contribution of a number of characteristics of the material of judgment to the final judgment. This method has been applied to both perceptual and abstract judgments.

Brunswik (*1944*) has applied the correlation method to the old psychophysical problem of size constancy. He studied a presumably representative sample of objects encountered by one subject as she went about her daily routine, having her judge the size of each object under different instructions. If we take as our problem the determination of the judgments of the size of the objects, the correlation coefficients supply valuable information. For Brunswik's one subject the net relationship between estimates of size and actual or "geographic" size was .98, while that between estimates and size of the retinal projections was .40. From these data it appears that perceptual judgments are more closely related to distal stimuli (objects in the environment) than to the proximal stimuli (retinal patterns) that mediate the judgments.

The correlation method has also been used for a similar analysis of judgments of prestige. Osgood and Stagner (*1941*) obtained ratings of the prestige of 15 common occupations and also ratings of the hopefulness, honesty, idealism, hours of work, and the like for these occupations. They then correlated the ratings of these qualities with the ratings of prestige and found, for example, that the ratings of hopefulness correlated .99 with prestige, while the ratings of hours

of work correlated only .20. Taking these correlations as evidence of the specific factors toward which the judges directed their attention when they made their general judgments, one may conclude that "qualities like brains, leadership and self-assurance are shown as marks of prestige," while "honesty, idealism, and congeniality are not."

The same technique has been applied to judgments of prominent psychologists. Fifteen graduate students of psychology were asked to rate 27 well-known psychologists, living and dead, in respect to (1) "original independent research," (2) "original theoretical ideas," (3) "teaching and stimulation of contributions by others: intellectual leadership," (4) "integration of facts and principles: systematizing," and (5) "overall claim to fame." The problem is to determine which of the first four factors has the most influence on judgments of "overall claim to fame." The correlations were, in the order listed, .52, .82, .91, and .75. Evidently these students did not give much weight to original research but weighted intellectual leadership most heavily.

The ideal of this type of analysis—though no one has carried it that far—appears to be a multiple regression equation, identifying and weighting all the variables that have significant effects on the final overall judgment. The variables must of course be known to the experimenter and the subjects in order that they can be correlated with the overall judgment.

Identification and Weighting by Factor Analysis. It is possible that a better analysis of what people are judging can be obtained by factor analysis, which would yield new hypothetical stimulus variables not necessarily recognized as such by the judges. This approach is particularly promising when complex objects are judged affectively. What are the variables, for example, to which people are responding when they judge some pictures *beautiful,* others *moderately appealing,* and others *unpleasant?*

A general factor of some importance is usually found in affective judgments when the technique is aimed in that direction: in judgments of odors (Beebe-Center, *1932;* M. D. Eysenck, *1944*), pictorial art (H. J. Eysenck, *1940*), and jokes (H. J. Eysenck, *1942*). After the general factor is removed, bipolar factors are often disclosed. For example, in Eysenck's (*1940*) study of judgments of portraits, bookbindings, vases, and the like the formal-representative polarity made an appreciable contribution to the overall aesthetic judgments. His study of humorous material (*1942*) indicated that the judgments were

influenced by the personal-impersonal, simple-complex, and sexual-nonsexual dimensions. A study of responses to jokes by Cattell and Luborsky (*1947*) resulted in clusters with such names as "debonair sexuality," "derision-superiority," and "anti-authoritarian." Following the same strategy, Bendig (*1952*) analyzed the interests of psychologists by having them judge the merits of psychological journals.

No one believes that the influence of the stimulus variables on the judgment is completely described by correlation coefficients or factor weights. The limitations of the correlation technique are well known today, and long ago Thorndike (*1918*), in a paper called "Fundamental Theorems in Judging Men," was calling attention to nonlinear relations between the independent variables and the judgment. The competent impressionistic or intuitive judge of men, he wrote, takes these complexities into account, hence the strength of the intuitive judgment, as compared with the more formal types, is that it is more quantitative! Despite these limitations the correlation technique has a definite value if it leads to identification and approximate weighting of the more important stimulus variables. This narrows the range of subsequent inquiry into the operations that the judge performs upon these stimulus variables.

ORGANIZING THE STIMULUS MATERIAL

Identification of the stimulus variables that influence the judgment and estimation of the strength of each does not tell the whole story. The next question is one of organization. All kinds of stimulus variables influence the judgment, but the judgment itself is a brief integrated response. What operations does the judge perform on this diversified stimulus material to get it into shape for judgment?

There are, of course, many patterns of organization of the data of judgment, many ways of structuring the stimulus material. These types of structure are discontinuous and cannot conveniently be ordered in a systematic fashion. But a few of the common organizing operations can be described, and the conditions that lead to the use of one rather than another.

A SINGLE STIMULUS DIMENSION AS FIGURE

Though we may speak of judgment as directed toward a simple dimension such as length, and we may find the responses system-

atically related to measured variations in length, this relationship does not tell us all that the judge did. Even in the simplest, best-controlled psychophysical experiment, in which the experimenter makes the subject's task as easy as possible, before the subject judges length he organizes the stimulus situation in such a way that length is abstracted out as figure against a variegated background of tactile, auditory, visual, affective, temporal, and other features. This figure-and-ground organization is a consequence of the subject's preparatory set, adopted at the experimenter's request, and of the experimental arrangements. The experimenter arranges the conditions so that when the subject compares a variable line with a standard line, the two lines are alike in all respects but length. And usually he begins with a large difference in length, so that the length aspect of the situation will be prominent. These two independent variables complement each other. That is, when the stimulus difference is small, the preparation is highly important. And when the preparation is inadequate, the stimulus difference in length must be large lest some other aspect of the stimulus situation become the figure.

Obviously experience plays a part in the organization of the stimulus material. The inspector in a factory, for example, has learned from experience what defects to look for and what deviations can be tolerated in judging the acceptability of the product. That aspect of the product in which he has found defects most frequently probably stands out as figure to him. The organization may change, of course, and some other aspect may become figure for the same reason. The employer judging an applicant for a job furnishes another example if experience has taught him to concentrate on a few highly desirable qualifications. Manifestations of these qualifications are noted while the rest of the applicant's behavior remains in the background.

When the stimulus objects vary in several characteristics, with no one characteristic standing out prominently, it is hard to abstract any one of these out of the complex. More is therefore required of the judge in the way of a preparatory set. He has to have good control of his thought processes to be sure that he is judging what he is prepared to judge and not something else. One meaning of "good judgment" could be the ability to judge different characteristics of a group of objects when prepared by different instructions. That is, he should be able to rate these objects at one time for weight, at another time for size, and again for attractiveness, and so on. Presumably the better his

judgment the less one judgment is confused with another. If the stimulus dimensions are objectively independent, "good judgment" would be indicated by a low correlation between judgments on these different dimensions. The effects of various conditions on ability to maintain set has come in for some study (see page 108), but apparently individual differences in this ability have not been seriously investigated.

ALGEBRAIC ADDITION

Suppose the first primitive organization of the stimulus situation has been accomplished so that one stimulus dimension, e.g., length, stands out as figure. Judgment of any stimulus object in respect to this dimension, which would be called a simple judgment, may now be assumed to be a resultant of opposing tendencies. If the response is not immediate, that is, if the process is one of judgment rather than perception, tendencies to call it *long* and *short* are both aroused, and the overt response is the algebraic sum of the two. This assumption underlies most quantitative theories of judgment (Boring, *1917;* Landahl, *1938;* Cartwright and Festinger, *1943;* Johnson, *1944*).

When complex judgments are made, it is possible that the judge simply adds the effects of all pertinent variables, whether positive or negative. If a person is debating a course of action, for example, he will often consider the matter from several angles, weighing advantages and disadvantages. He may even follow the advice of the judicious Poor Richard and list the advantages on one side of the page and the disadvantages on the other, then combine these in his final summing up. This is the model that is assumed, whether explicitly or not, in the studies of identification and weighting summarized above. As noted above, this is a good first approximation, whether it holds in all details or not.

This assumption of additivity should be critically examined. It is not just a theoretical question, for a different assumption is often made in the conduct of practical business enterprises. Some employers assert that an employee will be happier with $4000 a year and a $200 bonus than with a straight salary of $4200. The widespread use of premiums and prizes by merchandisers assumes that the customer would rather pay X dollars for a given quantity of coffee and get a "plus value" in the form of a teaspoon worth A dollars than pay $X - A$ dollars for the coffee. Mathematically, the merchandisers' assumption may be

written: $X - A + A > X$. The additivity assumption is written: $X - A + A = X$.

There have been only a few attempts to test the assumption of additivity. One procedure has been to get ratings of two objects of judgment, e.g., color cards, separately, then to get a rating of the combination, and to correlate series of such ratings. Washburn, Haight, and Regensburg (*1921*), for example, obtained a correlation of .75 between ratings, on a seven-point scale, of the pleasantness of 32 combinations of two colors and the sum of the separate ratings of each member of the combinations. Surely the judgments of the combinations must have been influenced largely by the pleasantness of the separate colors. But 15 percent of the disagreeable colors made pairs that were judged agreeable. Hence something more than simple addition must take place in some cases at least. Beebe-Center (*1932*) has summarized the evidence under the heading of "hedonic summation." In general it appears that when the separate stimuli are heterogeneous, as when the object is judged separately for form and color, simple algebraic summation is likely to occur; but in some cases the result of juxtaposing the separate stimuli may be not merely a combination, but a new configuration.

A different kind of evidence on this hypothesis may be obtained from an analysis of data reported by Thornton (*1943*); see page 296. He had large groups of judges rate photographs of men and women with and without glasses, smiling and serious. They rated them for six traits on an 11-point scale. It is possible to test the assumption of additivity by rearranging Thornton's data as in Table 9.1.

In this table "increment due to glasses" was obtained by subtracting the mean rating of the photographs without glasses or smile from mean rating of photographs with glasses. "Increment due to smile" was obtained by the analogous subtraction. The theoretical rating of photographs with glasses and smile was computed by algebraic addition of the values in the three lines above. The entries in the fifth row were obtained empirically. The discrepancies in the bottom row do look rather small in comparison to the variation within the table.

These experiments do not yield any clear-cut conclusion on additivity, but they permit the statement that under some conditions the judgment of a complex object approximates the sum of the judgments of its parts. When this principle does not hold, the reason may be that

TABLE 9.1. Analysis of Mean Ratings of Photographs for Six Traits Under Four Conditions.[a]

	Kind-liness	Intelli-gence	Industri-ousness	Hon-esty	Depend-ability	Sense of humor
Basal condition: no glasses, no smile	5.39	6.40	6.00	6.11	5.75	4.77
Increment due to glasses	.40	.82	1.07	.85	1.10	−.64
Increment due to smile	1.26	−.61	−.04	.48	.25	2.16
Glasses and smile, theoretical	7.05	6.61	7.03	7.44	7.10	6.29
Glasses and smile, obtained	7.16	6.58	6.75	7.33	6.92	6.70
Discrepancy	.11	−.03	.28	−.11	−.18	.41

[a] Data from Thornton *(1943)*.

the parts combine to form a new configuration which is not the sum of its parts.

We have to admit that the conditions under which the additivity principle does and does not hold have not been well described. Attempts to determine these conditions will be complicated by three considerations. (1) Errors in predictions based on the additivity principle may be due not to failure of the principle but to the use of scales of unequal units. Modern methods of scaling should be able to overcome this difficulty. (2) The errors in judgment data are usually too large to permit testing a precise hypothesis. One way to overcome this difficulty is to compare one hypothesis with another and to arrange the conditions so that the difference between the predictions from the two are large relative to errors of measurement. (3) The hypothesis of algebraic addition is a precise hypothesis, and as such can be disproved. Disproof of this hypothesis, however, does not prove any other hypothesis. Therefore the present need is to describe other, nonadditive ways of organizing the material of judgment.

SERIAL JUDGMENT OF HETEROGENEOUS ITEMS

What makes some judgments difficult is that there is too much to think about, too much data to organize. The judge may consider each item serially, not adding pros and cons, but using each item in differ-

ent ways to aid the overall decision. At least two such types of organization can be described.

THE CHECK LIST. One of the commonest judgment problems, one on which a man's reputation for "good judgment" often depends, is the decision of readiness for a critical activity. An airplane pilot, for example, has a large number of instruments to check before deciding that he is ready to take off. One practical way of handling this problem of quantity has been in use for many years, namely, the check list. The pilot memorizes a list of items to check before the critical take-off, and another to check before the critical landing. He does not balance positive variations in one item against negative variations in another. Each item makes its peculiar contribution to his decision.

SUCCESSIVE HURDLES. Consider the case of a grocer debating expansion of his store. He may first raise the most critical question: Can I really do more business? If the answer is No, the matter is closed. If the answer is Yes, he goes on to the next hurdle: Can I get more space? If the answer is No, the matter is closed. If the first two hurdles are surmounted, he goes on to the next: Can I finance the addition? And so on.

There is no way of knowing how frequently this procedure is followed, but it is likely that many decisions that seem superficially to be a result of algebraic addition are really reached by successive hurdles, or some mixture of the two. For the special case of the selection of personnel, the procedure has been discussed and recommended by Toops (*1945*).

JUDGMENT BY GENERAL IMPRESSION

Several kinds of evidence suggest that the various aspects of a complex object of judgment are not always judged separately for what they are worth—even though the instructions may call for separate judgments. The judge may get a general impression of the object of judgment, perhaps from one striking aspect, and judge the object predominantly on that basis, either ignoring the other aspects or treating them as minor deviations from the general impression. As noted in the first of the principles listed above, the importance of the general impression may not be recognized by the subject.

When the object of judgment is a person and the person is to be judged or rated as to several aspects, such as sociability, dependability, intelligence, and the like, it usually turns out that the person is rated

high on all traits, or low or medium on all traits. Thus when many people are rated on several traits, the trait ratings will be positively correlated. At first glance this implies judgment on the basis of a general impression, as contrasted with independent judgment of the separate traits. The phenomenon has in fact been called the "halo effect" (Thorndike, *1920*) and attributed to failure to analyze the complex object into its separate aspects. Despite the popularity of this effulgent phrase, when a search was made for evidence of the "halo effect," none could be found (Johnson, *1945*).

It is true that ratings of personality traits are usually correlated, but we do not know whether the correlation is due to judgment on the basis of a general impression or to an objective correlation between the traits. If we knew that the traits were actually independent, we could assume that the correlation is produced by the judging process, i.e., a halo effect. Or, if we knew that the judgments of the traits were made independently, we could assume that the correlations are in the people, as is assumed in factor analysis of trait ratings. But we must know one in order to infer the other. As it is, the facts referred to by the phrase "halo effect" are ambiguous.

Convincing evidence for the halo effect could perhaps be obtained by experimental manipulation of the judging process. On one occasion the people could be rated on one trait, then on another, and so on, one trait at a time. On another occasion one person could be rated on all traits, then another person on all traits, and so on, one person at a time. If the second procedure, which presumably gives less encouragement to analysis and judgment of the separate aspects, yields higher correlations than the first, the difference must be in the judging process and could very well be called a halo effect.

Sells's (*1936*) demonstration of the atmosphere effect in judgment of syllogisms illustrates a kind of judgment by general impression. Negative premises establish a negative atmosphere. Positive premises establish a positive atmosphere. Such unanalyzed general impressions or atmospheres influence judgment of the conclusion of the syllogism along with logical relationships, but syllogisms can be constructed in which the two influences are independent. When this was done and the influence of logical relationships balanced out, predictions on the basis of atmosphere effect were confirmed (see page 245). Methodologically it is noteworthy that the syllogisms were constructed by the experimenter and the two variables of atmosphere effect and logical

relationships experimentally controlled. It is not possible to test for the halo effect in this way because people cannot be experimentally constructed so that the traits are objectively independent.

By manipulating the procedure of judgment Asch, Block, and Hertzman (*1938*) were able to show that judgments of occupations may be organized around a predominant characteristic or general impression. In one part of their study they had college students rank ten professions according to six characteristics, such as intelligence and social usefulness. As one would expect from the above discussion, the rankings of the different aspects of these complex objects were correlated. But of course we cannot say at this point whether this correlation is a property of the judging process, i.e., a halo effect, or a property of the objects of judgment. Next the experimenters manipulated the judging process, without changing the objects of judgment. They introduced a group or majority standard similar to the subjects' own standards by printing on the blanks the mean rankings for intelligence obtained from a similar sample of students. This operated to raise the correlations slightly, presumably by strengthening the subjects' general impression of each occupation. Next a fictitious standard was introduced, one which was very different from the mean rankings of the previous sample. The correlations produced under the influence of this unconvincing standard were definitely lower and some were negative, presumably because the influence of the predominant characteristic had been destroyed. Further studies with other kinds of material, such as photographs, names of political figures, and slogans, yielded rather clear-cut evidence for the importance of a general impression or attitude toward each occupation underlying judgment of specific aspects.

When ratings of people are being obtained for practical purposes, as in selecting and promoting personnel, it is often assumed that a halo effect would be inaccurate; that it is better to judge people on several traits separately, then to combine the separate trait ratings statistically. Bingham (*1939*) has argued, on the contrary, that for most practical purposes the general overall impression of an employee's performance is just what is required, and that a direct unanalytical judgment of "usefulness to the company," or something similar, would be more valid than a combination of several analytical judgments.

The answer to this question, like the question about the existence of a halo effect, will be reached, not by statistical manipulation of data obtained by either judging procedure alone, but by comparing the

validity of judgments obtained by the analytical procedure with those obtained by the global procedure. One comparison of this kind has been reported by Coward (*1952*). A hundred essay examinations in English composition were graded on seven variables, such as material and organization, then these scores were weighted and summed to get analytical or "atomistic" grades. In contrast, "wholistic" judgments were also made. The results showed more consistency between analytical judges than between wholistic judges. The analytical judgments consumed more time.

COMPARING COGNITIVE PATTERNS

Up to this point we have found it profitable to consider separate aspects of a complex object of judgment, how and to what extent they are analyzed out of the stimulus complex, and how, in recombination, they interact to influence the judgment. But the judgment may be arrived at not so much by analysis as by manipulation and comparison of stimulus patterns as patterns. Certainly perceptual and affective judgments are often based on a comparison of patterns, but patterns of abstract material are more varied and permit more varieties of manipulation.

Suppose, for example, that a student finds a strange crawling animal under a rock, captures it, and brings it to a zoölogist for identification. The expert's task is to fit the specimen in front of him—not merely some aspects of it but the whole organism without remainder—into his classification system. The classification system is not up for judgment; but the specimen at hand is. The expert will examine separately various aspects of the animal, its legs, its coloring, its locomotion, and so on; but his final judgment depends on the pattern of all these, not on any algebraic addition of separate judgments. We may say in general that he matches the pattern facing him with the pattern of his taxonomic system to see where, in his judgment, it fits.

The physician examining a patient in a coma makes his judgment by a similar organization of the facts that he observes. He knows the pattern of symptoms of head injury, of diabetic coma, of alcoholic intoxication, and other possibilities. If any one aspect of the patient's condition suggests head injury, he checks for other corroborative signs of head injury. Before making his final decision he assembles all his information into a pattern and matches this pattern with the familiar symptom patterns.

What is sometimes called logical reasoning may proceed in the same way. The thinker tries to structure the facts of the case into a familiar form, perhaps into the form of a syllogism. The first judgment, then, is whether the data do fit into a familiar pattern. If so, the remainder of the logic is likely to be automatic. Reasoning in this instance is more an abstract judgment than a complicated process of problem solving.

Judgments of this sort are facilitated, of course, by mastery of standard or ideal patterns to which present patterns are compared. The zoölogist knows his species types. The physician knows his symptom patterns even if they are only textbook cases. The logician knows his logical types. The following quotation from a textbook on judging dairy cattle (Harrison, *1940*) shows the importance of ideal types in the livestock pavilion.

> The purpose of the true type cow, as adopted by the various breed associations, is to set forth a standard that can be used as the basis of judging. The successful judge actually compares each cow that he studies with the ideal or true type cow, and selects the cow that most nearly approaches this perfection. It is highly important, therefore, that a student of judging study the true type model so as to acquire a knowledge of the type that constitutes perfection in a dairy cow.

The formation of patterns of abstract data, as of perceptual data, depends on both the objective data and the thinker's set. If the data, when first grasped, do not fit into a familiar or at least a sensible pattern, the thinker will probably restructure the data, in accordance with his set, into some alternative pattern. That is, if he is determined to use syllogistic reasoning, he will reorganize the data into a syllogistic pattern. The physician, if he is cautious, will not stop with his first pattern of the symptoms, but will try out other ways of organizing the facts. The same facts make different contributions to the judgment when organized as parts of different patterns. It is possible, by suitable manipulation of instructions, to study the processes by which people synthesize information and draw conclusions about other people (Dailey, *1952*).

The acquisition of abstract patterns and the production of new ones have been discussed in earlier chapters. The intellectual activity may be described as judgment, according to the present system of concepts, when patterns—otherwise produced—are compared and the response is limited to a few categories, such as *same, different,* and *Type A.*

CONTEXT OF JUDGMENT OR FRAME OF REFERENCE. It is not only the object of judgment which determines the final decision, but also the background pattern or context within which the object is judged. The unknown bug is judged by the zoölogist, not in isolation, but in relation to his classification system. The patient is judged by the physician in relation to a system of diagnostic categories. Because of the figure-ground effect the background pattern or context, within which the object of judgment is evaluated, is often not scrutinized but taken for granted.

Consider a case at law as an illustration. A man is involved in a fight on the street. A policeman accuses him of disturbing the peace and brings him before a judge. The judge's examination of the accused and the witnesses will bring forth many data: his name, his occupation, perhaps his reputation, several descriptions of the alleged offense, the time of day, the events leading up to the offense, and so on. Obviously these facts of the case will not be summated algebraically. Some of the facts, meanings, implications, connotations, associations, suggestions, etc., admitted to judgment will be organized into a description of the alleged offense, i.e., the object of judgment. And some of the data will be organized, probably less explicitly, into a context or frame of reference within which the act is judged.

The term frame of reference, which has achieved some popularity in recent psychological writing, has an important place in the psychology of judgment. But, like many valuable terms, it is in danger of being used so loosely that its value is dissipated. Any new fact brought to judgment is sometimes said to change the frame of reference. But surely this is too loose a meaning. Buxton (*1942*) has defined the frame of reference as "the background of stimulation which influences our behavior in a particular situation." The word background implies that something is in the foreground, obviously the object of perception or judgment, the stimulus pattern to which the organism is set to respond. The frame of reference exerts its influence by supplying background for the object of perception or judgment. In the illustration above, if the judge brings out the fact that the accused used both fists and blacked both the victim's eyes, this new information is not part of the frame of reference but is part of the object of judgment itself. The frame of reference is changed if the court discovers that the victim had stolen the accused's wallet or had called him vile names.

This term is one which originated in the study of perception and

was extended to the study of judgment. As applied to judgment, the term context is more appropriate because it is less restricted in meaning, less of a metaphor when applied to nonspatial objects of judgment, but the two seem to be used interchangeably. Illustrations of the influence of a frame of reference, and of changes in a frame of reference, on perception are presented in Chapter 4. The background of judgment differs, however, from the background of perception in that the latter is organized quickly, by direct perception of the stimulus field; while the context of judgment is organized more slowly from the facts at hand, from memories, associations, etc., by abstract or conceptual operations. Just as painters, architects, composers, photographers, camoufleurs, illusionists, and experimental psychologists have learned how to manipulate the background of perception, so have public relations counselors, debaters, attorneys at law, political campaign managers, and experimental psychologists learned how to manipulate the frame of reference, each for his own ends. And in a court of law the accused may even claim that he has been "framed."

Judgments of oneself illustrate the importance of the context of judgment, especially estimates of one's achievement or one's aspirations for future achievement. Many factors enter into a judgment of one's achievement, one of the principal factors being, as would be expected, past success and failure. Any one present achievement and any one aspiration for the future is likely to be judged against a context organized out of remembered past achievement.

Often, also, the context includes the achievement of others. This can be demonstrated rather easily in the laboratory because the experimenter can control the information about others' achievement which he gives to the subject (Chapman and Volkmann, *1939;* Gould and Lewis, *1940;* Hertzman and Festinger, *1940*). If, for example, a person who makes a score of 40 on a vocabulary test is told that the average professional writer makes a score of 38, he will probably accept his achievement as pretty good. If, however, he is told that fifth-grade children make an average score of 41, he will rate himself low, and aim higher. Obviously the difference between these two instances is not in the object of judgment, which remains the achievement of a score of 40, but in the background or context or frame of reference within which it is judged.

When the performance judged is the performance of another person, it also is judged within a context organized out of information from

many sources. The practice of computing norms for the interpretation of the achievement shown by test scores is in fact an application of this principle. To put it psychologically we may say that the function of test norms is to supply the users of the test results with a standard frame of reference.

The opinions of others can also operate as context. Asch (*1940*) had two samples of students from the same college population rank ten occupations in respect to intelligence, social usefulness, and the like. Then they were given, "as an illustration," the ranking of a particular occupation, politics, by a group of 500 college students. For one group this fictitious ranking put politics on top, for the other politics was on the bottom, and the resulting rankings from the students showed the usual shift in the direction of the majority opinion. But when Asch asked the students what politicians they had in mind while making the judgments, the first group used phrases like "more statesmanship than politics," while the second group spoke of the "usual neighborhood politicians." The difference in the ranking of politicians by the two groups can be interpreted, therefore, as a result of the context introduced by the fictitious rankings: politicians are judged differently when they are on the bottom and when they are on the top.

A similar experiment by Lewis (*1941*) asked for judgments of slogans, such as "Balance the Budget." For different groups the slogans were attributed to Herbert Hoover, Franklin D. Roosevelt, and Earl Browder, and the prestige of these experts exerted some influence on the subjects according to their political allegiances. But when the subjects were asked what they thought of the expert's rankings, their answers indicated that many subjects did not accept the suggestions blindly, but rather tried seriously to compare their interpretation of the slogan with their knowledge of the expert and to make sense of any apparent contradictions. According to Lewis, the source of the statement supplies context for judgment of the slogan, and a change of opinion depends, not on blind acceptance of suggestion, but on whether the new pattern created seems as clear and correct as the old.

In these experiments by Asch and Lewis we see that shifts in judgment usually attributed to the influence of majority and expert opinion may equally well be attributed to the effect of context in encouraging an intellectual reorganization of cognitive patterns. (Also see Asch, *1948*.) This statement may be reversed. When enthusiastic teachers of

art and literature attribute shifts in their students' taste to intellectual appreciation, critics may claim that the shifts are due to the attachment of prestige to certain names. It is hard to separate these two principles of judgment and weight their relative strengths. Comments of the judges cannot be taken as critical evidence unless such comments can be shown to have high internal reliability or to lead to good predictions. Usually both principles operate simultaneously, though cognitive reorganization will be found more often in the more intelligent thinkers and suggestion more often in children and adults of below-average intelligence.

There are many ways of structuring and restructuring the materials of judgment. (For a comprehensive list of suggestions, see Blumenfeld, *1931,* pages 184–188.) When the thinker patterns the available data and fits one pattern into another and soon makes his response, we may call it judgment. But often a productive process may intervene, as when the thinker invents a new way of organizing the data, or thinks of new implications of the data. When this takes place, a productive process is added to the judgment process.

INDIVIDUAL DIFFERENCES

In all the experiments mentioned in this chapter individual differences in judgment have been observed. But some of these are more promising for differential psychology than others.

Individual differences in susceptibility to suggestion have been studied for many years. The scores are rather specific to the kind of material used and the source of the suggestion; some small degree of generality has been found. (See Eysenck and Furneaux, *1945,* for references.) The same sort of variation in distortion of judgment by affects and motives has been noticed, but scores of this kind have not been analyzed as measures of individual differences.

The possibilities of this approach are illustrated by a study of humor by Luborsky and Cattell (*1947*). Having previously (see page 304) arranged a large number of jokes in clusters, they asked college students to judge the merit of each joke. These students also took the Guilford-Martin test of personality. A few of the 13 humor clusters correlated with a few of the 10 personality factors. For example, students with high scores on personality factor R, "a tendency to a happy-go-lucky, care-free disposition," gave high ratings to jokes in the cluster called "debonair sexuality." Students who got high scores on

personality factor I, "self-confidence and lack of inferiority feelings," gave favorable judgments of jokes in the cluster called "sturdy irony and dominance." These correlations, as far as they go, give some clue to individual differences in the motivation of judgment and, in reverse, give some encouragement to the use of judgments of jokes as an objective test of personality.

When a number of judges rate complex objects in respect to several stimulus variables, a variety of interesting and promising scores may be obtained from each subject.

1. The correlation of each judge's ratings with the ratings of the group as a whole may be taken as a measure of his agreement with the group. Guilford and Holley (*1949*) have recommended a measure of this kind as an indication of objectivity of judgment.

2. If we can assume that the variables rated are independent, the mean of the correlations between the ratings of any judge can be taken as a measure of his vulnerability to the halo effect. Or this coefficient subtracted from unity is a measure of the judge's ability to rate one thing at a time, his analytical ability. If we cannot assume that the variables are independent, we cannot say what proportion of this coefficient is halo effect; but if we can assume that the variables are objectively the same for all judges, comparisons between values of this coefficient for different judges are meaningful (see Crissy and Regan, *1951*).

3. If overall or summary ratings of complex objects, e.g., prominent psychologists, are made as well as ratings of specific characteristics, e.g., research contributions and intellectual leadership, the correlation of each series of specific ratings with the overall ratings can be computed for each judge. These correlations indicate how much each judge is influenced by each variable in making his overall judgment. We may find that research contribution is the most important factor for some judges, while other judges put most weight on intellectual leadership.

4. If the judgments are repeated, a measure of stability or repeat reliability can be obtained. Internal consistency at any one time can also be computed in the usual way.

"Good Judgment." This term that we hear so often is usually used in a rather vague way. It could refer to consistency or stability of judgment, as in the fourth suggestion above. It could refer to analytical ability, as in the second suggestion above. The procedure used by Postman and Page (see page 108) offers another possibility. People can be

asked to judge objects in one respect, then in another respect, then again in the first respect. The drop in accuracy as a result of interpolated activity is a measure of susceptibility to interference or lack of analytical ability. It might be correlated with the second suggested coefficient above. Another approach to "good judgment" will be described in Chapter 11.

Whether any of these suggested measures of individual differences can be obtained with sufficient reliability for practical purposes is an open question. It is a question also whether any of these measures can be related to any criterion of success in practical pursuits outside the laboratory. But we know that many occupations call for judgments of one kind or another all day long. And the term "good judgment" is often emphasized by employers and job analysts. Many psychological processes are quite different in the laboratory and in real life, but a priori the judgment process seems to be very similar wherever it occurs.

REFERENCES

Asch, S. E., 1940. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards, *J. Social Psychol., 12*:433–465.

Asch, S. E., 1948. The doctrine of suggestion, prestige, and imitation in social psychology, *Psychol. Rev., 55*:250–276.

Asch, S. E., 1951. Effects of group pressure upon the modification and distortion of judgments, in *Groups, leadership and men,* ed. H. Guetzkow, Pittsburgh, Carnegie Press.

Asch, S. E., Block, H., and Hertzman, M., 1938. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment, *J. Psychol., 5*:219–251.

Baker, L. E., 1937. The influence of subliminal stimuli upon verbal behavior, *J. Exper. Psychol., 20*:84–100.

Beebe-Center, J. G., 1932. *Pleasantness and unpleasantness,* New York, D. Van Nostrand Co.

Bendig, A. W., 1952. A Q-technique study of the professional interests of psychologists, *J. Psychol., 33*:57–64.

Berenda, R., 1950. *The influence of the group on the judgments of children,* New York, King's Crown Press.

Bingham, W. V., 1939. Halo, invalid and valid, *J. Appl. Psychol., 23*:221–228.

Birch, H. G., 1945. The effect of socially disapproved labeling upon a well-structured attitude, *J. Abnorm. & Social Psychol., 40*:301–310.

Blumenfeld, W., 1931. Urteil und Beurteilung, *Arch. f. d. ges. Psychol.*, Ergänzungsband 3.

Boring, E. G., 1917. A chart of the psychometric function, *Am. J. Psychol.*, *28*:465–470.

Brunswik, E., 1944. *Distal focussing of perception: size-constancy in a representative sample of situations,* Psychol. Monogr., No. 56.

Burtt, H. E., and Falkenburg, D. R., 1941. The influence of majority and expert opinion on religious attitudes, *J. Social Psychol.*, *14*:269–278.

Buxton, C., 1942. The frame of reference, in *Fields of psychology,* ed. R. H. Seashore, New York, Henry Holt and Co.

Cantril, H., 1938. The prediction of social events, *J. Abnorm. & Social Psychol.*, *33*:364–389.

Cantril, H., 1947. *Gauging public opinion,* Princeton, Princeton University Press.

Cartwright, D., 1941. Relation of decision-time to the categories of response, *Am. J. Psychol.*, *54*:174–196.

Cartwright, D., and Festinger, L., 1943. A quantitative theory of decision, *Psychol. Rev.*, *50*:595–621.

Cattell, R. B., and Luborsky, L. B., 1947. Personality factors in response to humor, *J. Abnorm. & Social Psychol.*, *42*:402–421.

Chant, S. N. F., and Salter, M. D., 1937. The measuring of attitudes toward war and the galvanic skin reflex, *J. Educ. Psychol.*, *28*:281–289.

Chapman, D. W., and Volkmann, J., 1939. A social determinant of the level of aspiration, *J. Abnorm. & Social Psychol.*, *34*:225–238.

Coffin, T. E., 1941. *Some conditions of suggestion and suggestibility,* Psychol. Monogr., 53, No. 241.

Coward, A. F., 1952. A comparison of two methods of grading English compositions, *J. Educ. Research, 46*:81–93.

Crissy, W. J. E., and Regan, J. J., 1951. Halo in the employment interview, *J. Appl. Psychol.*, *35*:338–341.

Cronbach, L. J., 1946. Response sets and test validity, *Educ. Psychol. Meas.*, *6*:475–494.

Cronbach, L. J., 1950. Further evidence on response sets and test design, *Educ. Psychol. Meas.*, *10*:3–31.

Cronbach, L. J., and Davis, B. M., 1944. Belief and desire in wartime, *J. Abnorm. & Social Psychol.*, *39*:446–458.

Dailey, C. A., 1952. The effects of premature conclusion upon the acquisition of understanding of a person, *J. Psychol.*, *33*:133–152.

Eysenck, H. J., 1940. The general factor in aesthetic judgments, *Brit. J. Psychol.*, *31*:94–102.

Eysenck, H. J., 1942. The appreciation of humor: an experimental and theoretical study, *Brit. J. Psychol.*, *32*:295–309.

Eysenck, H. J., and Furneaux, W. D., 1945. Primary and secondary sug-

gestibility: an experimental and statistical study, *J. Exper. Psychol.*, *35*:485–503.

Eysenck, M. D., 1944. An experimental and statistical study of olfactory preferences, *J. Exper. Psychol.*, *34*:246–252.

Farnsworth, P. R., 1945. Further data on the obtaining of Thurstone scale values, *J. Psychol.*, *19*:69–73.

Ferguson, L. W., 1935. The influence of individual attitudes on construction of an attitude scale, *J. Social Psychol.*, *6*:115–117.

Ferguson, L. W., 1940. Comparison of scale values from the method of equal appearing intervals and paired comparisons method, *J. Gen. Psychol.*, *23*:431–435.

Festinger, L., 1943. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference, *J. Exper. Psychol.*, *32*:291–306.

Fullerton, G. S., and Cattell, J. McK., 1892. On the perception of small differences, *Publ. Univ. Penn., Philos. Series, No. 2;* also in *Man of Science,* by James McKeen Cattell, Lancaster, Pa., Science Press, 1947.

Gould, R., and Lewis, H. B., 1940. An experimental investigation of changes in the meaning of level of aspiration, *J. Exper. Psychol.*, *27*:422–438.

Guilford, J. P., 1950. *Fundamental statistics in psychology and education,* New York, McGraw-Hill Book Co.

Guilford, J. P., and Holley, J. W., 1949. A factorial approach to the analysis of variances in esthetic judgments, *J. Exper. Psychol.*, *39*:208–218.

Harrison, E. S., 1940. *Judging dairy cattle,* New York, John Wiley & Sons.

Hartmann, G. W., 1936. A field experiment on the comparative effectiveness of "emotional" and "rational" political leaflets in determining election results, *J. Abnorm. & Social Psychol.*, *31*:99–114.

Hertzman, M., and Festinger, L., 1940. Shifts in explicit goals in a level of aspiration experiment, *J. Exper. Psychol.*, *27*:439–452.

Hinckley, E. D., 1932. The influence of individual opinion on construction of an attitude scale, *J. Social Psychol.*, *3*:283–296.

Hollingworth, H. L., 1925. The definition of judgment, *Psychol. Rev.*, *32*:337–361.

Hovland, C. I., and Mandell, W., 1952. An experimental comparison of conclusion-drawing by the communicator and by the audience, *J. Abnorm. & Social Psychol.*, *47*:581–588.

Hovland, C. I., and Sherif, M., 1952. Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales, *J. Abnorm. & Social Psychol.*, *47*:822–832.

Janis, I. L., and Frick, F., 1943. The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms, *J. Exper. Psychol.*, *33*:73–77.

Johnson, D. M., 1939. *Confidence and speed in the two-category judgment,* Arch. Psychol., No. 241.

Johnson, D. M., 1940. Confidence and the expression of opinion, *J. Social Psychol., 12:*213–220.

Johnson, D. M., 1944. Generalization of a scale of values by the averaging of practice effects, *J. Exper. Psychol., 34:*425–436.

Johnson, D. M., 1945. A systematic treatment of judgment, *Psychol. Bull., 42:*193–224.

Kelman, H. C., 1950. Effects of success and failure on "suggestibility" in the autokinetic situation, *J. Abnorm. & Social Psychol., 45:*267–285.

King, G. F., Ehrmann, J. C., and Johnson, D. M., 1952. Experimental analysis of the reliability of observations of social behavior, *J. Social Psychol., 35:*151–160.

Landahl, H. D., 1938. A contribution to the mathematical biophysics of psychophysical discrimination, *Psychometrika, 3:*107–125.

Lefford, A., 1946. The influence of emotional subject matter on logical reasoning, *J. Gen. Psychol., 34:*127–151.

Lewis, H. B., 1941. Studies in the principles of judgments and attitudes: IV. The operation of "prestige suggestion," *J. Social Psychol., 14:*229–256.

Lorge, I., and Curtis, C. C., 1936. Prestige-suggestion and attitudes, *J. Social Psychol., 7:*386–402.

Luborsky, L. B., and Cattell, R. B., 1947. The validation of personality factors in humor, *J. Personality, 15:*283–291.

Lund, F. H., 1925. The psychology of belief, *J. Abnorm. & Social Psychol., 20:*180–190.

Lurie, W., 1938. Measurement of prestige and prestige-suggestibility, *J. Social Psychol., 9:*219–225.

McGarvey, H. R., 1943. *Anchoring effects in the absolute judgment of verbal materials,* Arch. Psychol., No. 281.

McGregor, D. M., 1938. The major determinants of the prediction of social events, *J. Abnorm. & Social Psychol., 33:*179–204.

Marple, C. H., 1933. The comparative susceptibility of three age groups to the suggestion of group versus expert opinion, *J. Social Psychol., 4:*176–186.

Miller, J. G., 1939. Discrimination without awareness, *Am. J. Psychol., 52:*562–578.

Moore, H. T., 1921. The comparative influence of majority and expert opinion, *Am. J. Psychol., 32:*16–20.

Mosier, D. I., 1941. A psychometric study of meaning. *J. Soc. Psychol., 13:*123–140.

Osgood, C. E., and Stagner, R., 1941. Analysis of a prestige frame of reference by a gradient technique. *J. Appl. Psychol., 25,* 275–290.

Pintner, R., and Forlano, G., 1937. The influence of attitude upon scaling of attitude items. *J. Soc. Psychol. 8,* 39–45.

Saadi, M., and Farnsworth, P. R. 1934. Degrees of acceptance of dogmatic statements and preferences for their supposed makers. *J. Abn. Soc. Psychol. 29,* 143–150.

Saffir, M. A., 1937. A comparative study of scales constructed by three psychophysical methods. *Psychometrika, 2,* 170–198.

Sells, S. B., 1936. The atmosphere effect. Arch. Psychol. No. 200.

Sherif, M., 1936. The psychology of social norms, New York, Harper & Bros.

Thorndike, E. L., 1918. Fundamental theorems in judging men. *J. Appl. Psychol., 2,* 67–76.

Thornton, G. R., 1943. The effect upon judgments of personality traits of varying a single factor in a photograph. *J. Soc. Psychol., 18,* 127–148.

Thornton, G. R., 1944. The effect of wearing glasses upon judgments of personality traits of persons seen briefly, *J. Appl. Psychol., 28:*203–207.

Toops, H. A., 1945. Philosophy and practice of personnel selection, *Educ. Psychol. Meas., 5:*95–124.

Washburn, M. F., Haight, D., and Regensburg, J., 1921. The relation of the pleasantness of color combinations to that of colors seen singly, *Am. J. Psychol., 32:*145–146.

Williams, A., 1938. Perception of subliminal visual stimuli, *J. Psychol., 6:*187–199.

Witryol, S. L., 1950. *Age trends in children's evaluation of teacher-approved and teacher-disapproved behavior,* Genet. Psychol. Monogr., No. 41, 271–326.

Woodworth, R. S., and Sells, S. B., 1935. An atmosphere effect in formal syllogistic reasoning, *J. Exper. Psychol., 18:*451–460.

Chapter 10

Judgment: Response Variables

> A "heavy" baby is heavy in comparison with the general run of babies, but those other babies are not thought of while you are lifting the particular baby.
>
> —R. S. WOODWORTH

THE important dependent variables of judgment are (1) the response by which the judgment is expressed or communicated, usually called simply the judgment; (2) confidence or certainty of judgment; (3) the time taken by the judgment process; and (4) the difficulty of the judgment.

EXPRESSING THE JUDGMENT

Judgments are communicated most often by speech or by writing, less often by conventional gestures. Voting is done by pulling a designated lever in a voting machine, by raising the hand at a designated time, or by making a mark in a designated place on a ballot. The rules of many games specify the movements to be used to communicate a player's option or an official's decision, as when a baseball umpire raises one hand to indicate a strike and the other to indicate a ball.

In the laboratory the outcome of judgment is usually communicated from subject to experimenter by verbal expressions—such as *yes* or *no; lighter, equal,* or *heavier; pleasant, indifferent,* or *unpleasant*—or by checking corresponding boxes on a blank answer sheet, or by some other designated response, such as pushing a key.

PREPARATION FOR THE RESPONSE

By definition a judgment is not an automatic reflex. The judge performs some intellectual operations on the stimulus material that outlast the reaction time, and this judging activity is influenced by an identifiable preparation or set. The lost automobile driver, for example, may be set to choose one of a small number of environmental possibilities, perhaps two or three alternative routes. If he is set thusly, his preparation has eliminated cross-country travel, complete reversal, spending the night on the road, etc. In this example the thinker is set to express his judgment, not by a verbal response, but by a specified course of action. The stimulus side of the preparation, which directs his intellectual efforts toward some aspects of the situation rather than others, was discussed in the preceding chapter.

Similarly, if the response is to be a verbal one, as expressed by a judge in a court of law, a voter in a voting booth, a student taking a true-false test, or a laboratory subject in an experiment on judgment, the preparation specifies the categories in which the judgment is to be made. In these cases the preparation is established by rules of law, conventions, instructions from examiner or experimenter, or by one's specifications to himself of the requirements of the final judgment. That is why it is correct to say that *all judgments are relative* and can only be appreciated in relation to the other alternative judgments that the judge was prepared to make.

In laboratory experiments the set established by the experimenter's instructions soon becomes habitual, and if the subject returns to the experiment after a lapse of a week, he will adopt the old set without being told. When the judgment comes in the middle or at the conclusion of a complicated problem-solving enterprise, the set to respond in one category or another is part of the larger enterprise. The thinker, searching for the solution to a problem, who stops to test a promising hypothesis will judge that hypothesis in respect to its contribution to his overall activity. He may be set, then, to judge the hypothesis *acceptable* or *not acceptable*. The poet may evaluate his product as *complete* and quit work, or he may judge it *incomplete* and continue revising. In general the set that determines the form in which the judgment is expressed may be established in all the many ways that any other set may be established.

The responses by which the judgment is expressed are necessarily simple and few. In fact it is part of the definition of judgment that the

situation be left simpler or more determinate than it was found. Prior to comparative judgment of two lines, for example, the relation between the two is not definite. In that respect the situation is ambiguous or problematic. The result of the judgment is that one line is asserted to be longer than the other—or the two are judged equal. In that respect, then, the situation is now clarified, made more definite, easier to adjust to and operate in—whatever the long-term consequences may be. The immediate consequence of the judgment, the nature and extent of the temporary clarification, depends therefore upon the categories of response permitted by the preparation.

JUDGMENT IN TWO CATEGORIES. The simplest type of judgment, in respect to the response variables, is of course the two-category judgment. If only one category of response is permitted, the judge cannot exercise his judgment. If the ballot has a place only for *Jah,* or if it is dangerous to vote *Nyet,* the judge has no choice, and the procedure does not settle anything.

The two-category judgment is illustrated when a jury is instructed to bring in a verdict of *guilty* or *innocent,* when students are instructed to mark test items *true* or *false,* and when the subjects of a psychophysical experiment are instructed to judge weights *heavier* or *lighter* than a standard weight.

The significance of a response in either category, i.e., the meaning of the judgment, may be self-evident or the categories may be defined *ad hoc.* The court, for instance, will often tell the jury quite explicitly what *guilty* means, what *innocent* means, and the consequences of each decision. The instructions for a psychological test may leave *true* and *false* with their conventional meanings or may assign a special meaning to each. In any case these instructions, whether imposed by another or by the thinker himself, establish the preparatory set. The thinker is set to make one of these responses, as thus specified, to the material at hand. This is quite different from the set for word association in which the stimulus word is definite but the response is specified only very generally, in terms of a relationship to the stimulus word.

It is important to note that the categories of response in the two-category judgment are open at both ends (see Figure 10.1). The defendant cannot be too guilty to be judged *guilty* or too innocent to be judged *innocent.* The test item cannot be too true or too false. The extremes are unlimited; the only limit is at the boundary where the categories meet.

True	False

FIGURE 10.1. Judgment on a scale of two categories.

The instructions, by defining the response categories, can modify either category of response, moving the boundary between them to the left or the right. A stringent definition of *true* ("Don't call it true unless all of the statement is true beyond a shadow of a doubt") will reduce the number of *true* responses and thus move the boundary to the left. A stringent definition of *false* will do the opposite.

JUDGMENT IN SEVERAL HETEROGENEOUS CATEGORIES. The two-category judgment is of course a special case, less common than is usually

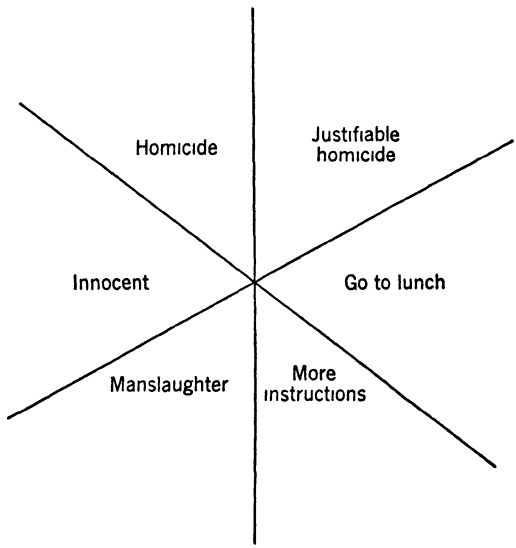


FIGURE 10.2. Judgment in several heterogeneous categories.

supposed. More often a variety of responses is permitted by the conditions of judgment. The lost automobile driver may, for example, have four possibilities of response: three possible routes and the possibility of waiting at the crossroad for someone who knows the roads. In a case at law the jury may bring in a verdict of homicide, of justifiable

homicide, of manslaughter, of innocence, or it may decide that it should ask the court for more instructions. Or it may decide to stop for lunch (see Figure 10.2). The number of alternative responses must be small and they must be fairly well specified, however, else the activity is not called judgment.

JUDGMENT IN SEVERAL RELATED CATEGORIES. Frequently the judgments are made in categories that are related to each other in some systematic way. The classification systems used by librarians in cataloguing books call for judgments of this kind, as do the taxonomic systems used by zoölogists. These might be called multidimensional judgments since the responses classify the objects of judgment in respect to several dimensions at once. Such multidimensional judgments have not been thoroughly studied by psychologists, but judgments that locate the objects of judgment along one dimension have been investigated in some detail. When the categories of judgment designate adjacent regions of a scale that is meaningful to the judge, tabulation of a series of judgments will usually show that the judge is making use of a response scale or scale of judgment. A large part of the psychology of judgment is concerned with the investigation of such response scales: their description, their organization, and their relation to other variables.

THE RESPONSE SCALE

When the attribute being judged varies along a continuum, e.g., a continuum of length, intelligence, sociability, or all-round merit, judgment may be expressed by responses that designate categories strung together along a scale. The significance of each category name for communicating the judgment may be explicitly and individually defined by the instructions, or the significance of some of them may be inferred or generalized from their relation to the others. In either case the judge is prepared to characterize the stimulus object by a category name indicating location on the scale. That is, the preparatory set restricts the judge's responses firstly and unequivocally to a few category names, and secondly and fallibly to that particular name that best expresses his judgment (see Figure 10.3).

Very unsociable	Unsociable	Sociable	Very Sociable

FIGURE 10.3. Judgment on a scale of four related categories.

The scale may be a well-known scale, familiar to many, as when judgments are made in inches or minutes, or when handwriting samples are judged as fifth-, sixth-, or seventh-grade quality. Or it may be peculiar to the individual judge, lasting perhaps only for the duration of one experiment, as when weights are judged as #1, #2, and #3 or personality variables are judged as —2, —1, 0, +1, and +2.

Scales of judgment, in contrast to contexts of judgment, are easy to work with empirically for they are simple and one-dimensional, while contexts may be multidimensional. Our definition of judgment implies that the response is more definite than the stimulus situation. But some confusion between the context of judgment and the response scale is possible. This can be clarified by taking up again our previous illustration of the man who was accused of disturbing the peace. Let us say that, within the context that the judge has constructed from the evidence presented and from his knowledge of law, he has judged the man guilty. Now he has to pronounce sentence, to communicate his decision to the accused and to society. His response will be limited by the rules of law, which state what he may or must or may not say. Within these limitations he will respond according to his location of the offense on his scale of the seriousness of offenses. He may consider it *very serious, serious, minor,* or *trivial,* and his decision must be communicated in such a way that the agents of the law can execute it. Thus we see that the context determines what independent variables or factors are considered in the judgment and what force they will have. The scale determines the category of response by which the judgment is expressed.

The context may be complex while the scale is simple. If we raise the question "Was President Franklin Roosevelt a great man?" the context is certainly complex, and indeed different people will use different contexts; but, if they follow instructions, they will all use the same two-category response scale made up of *Yes* and *No.* In certain special cases, as noted below, the context and the response scale are closely tied together, but for clarity of discussion the two must be conceptually distinguished. The object of judgment and its context are brought to judgment via the sense organs and memory. The response scale is a linear arrangement of alternative responses, whether the responses are verbal or manual. Both are influenced by the set and both influence the process of judgment, but context is a property of the stimulus material that frames the object of judgment while the re-

sponse scale is a property of the responses that scales the communicative significance of the judgment. Just as identification of stimulus variables and context defines the thinker's momentary effective environment, so definition of his response scale defines the significance of his behavior.

METHODS FOR DESCRIBING THE RESPONSE SCALE

What does it mean when a judge calls a stimulus object *long?* The communicative significance of such a response is partly determined in advance of its use in judgment by the conventional meaning of the response word. It is determined in advance also by the response words included in the response scale. That is, when the response scale includes the four terms, *very short, short, long,* and *very long,* the rank of each term is prepared in advance. We can be sure, for example, that *short* will be used to refer to lengths between *very short* and *long.* If we wish to go beyond this ranking and determine just what lengths are meant by each response of the scale, quantitative investigation is necessary.

Scaling techniques are often used to get logically defensible measures of psychological magnitudes. Though the computations are similar in some respects, our purpose here is quite different—because the logic of measurement is different from the psychology of judgment. Our purpose is to describe the communicative significance of the response scale that the judge actually uses, not to transform his judgments into measures that can be statistically manipulated. In general there are four methods: measurement of category limens, average stimulus in each category, reproduction methods, and rating methods.

Measurement of Category Limens. If the objects of judgment are weights, measured objectively in grams, and the judge is prepared to respond in two categories, *light* and *heavy,* the data will be frequency of use of each response category for each stimulus. Conventional psychophysical computations, as for the method of constant stimuli described by Woodworth (*1938*), yield a limen, or threshold or boundary, between these two categories which is measured in the units of the stimulus objects, e.g., in grams. This limen is a statistically determined transition point at which the probability of judgment in one category equals the probability of judgment in the other category (Rogers, *1941;* Cartwright, *1941b*). If the context of judgment includes a standard stimulus, the response categories are *lighter* and *heavier* and the limen

is then called the "point of subjective equality." If no standard is used, this point is merely a limen between adjacent regions of the response scale. The only important assumption made when computing a limen from a distribution of judgments in this way is that the distribution of transition points is fairly represented by an average point or limen.

The method is a general one that can be used to describe a scale of any number of categories in terms of the limens between any two adjacent categories. The number of category limens is one less than the number of categories. Wever and Zener (*1928*) described response scales of three categories. Blumenfeld (*1931*) described scales of two, three, five, and ten categories. When the category limens are computed, the width of each category in stimulus units is known, and these measurements supply a quantitative record of the judge's definition of each category of judgment. The width of a category may also be called a "range of equivalence" (Cartwright, *1941a*). The width of each end category is indeterminate because each is open at the extreme.

An experiment in judgment of lifted weights (Johnson, *1944*) will serve as an illustration. One subject was presented a wide range of

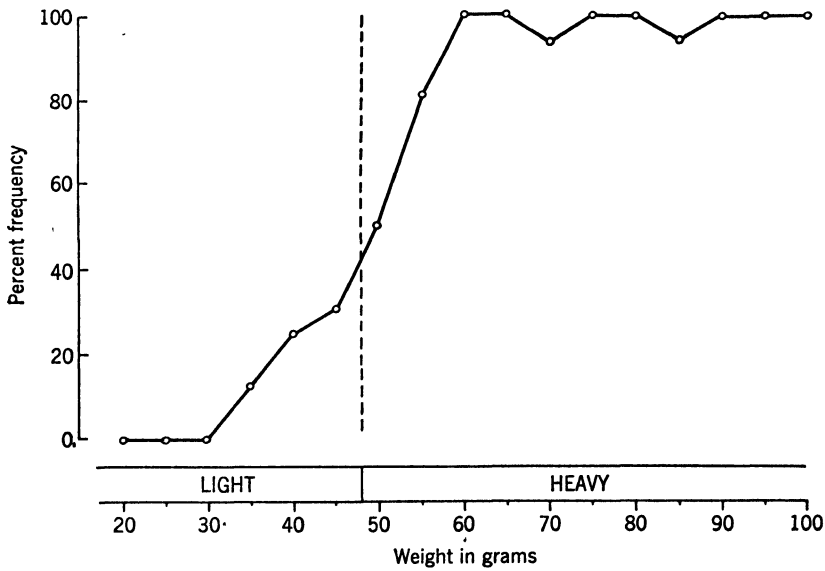


FIGURE 10.4. Transition zone and category limen for a two-category response scale. One subject judged each weight 16 times. From Johnson (*1944*).

weights one at a time and asked to judge each *light* or *heavy*. The weights ranged by steps of five grams from 20 to 100 grams. After a little practice those above 48 grams were called *heavy* more often than not, and those below 48 grams were called light. This subject's response scale for this series of judgments can therefore be represented, as in Figure 10.4, with a category limen at 48 grams.

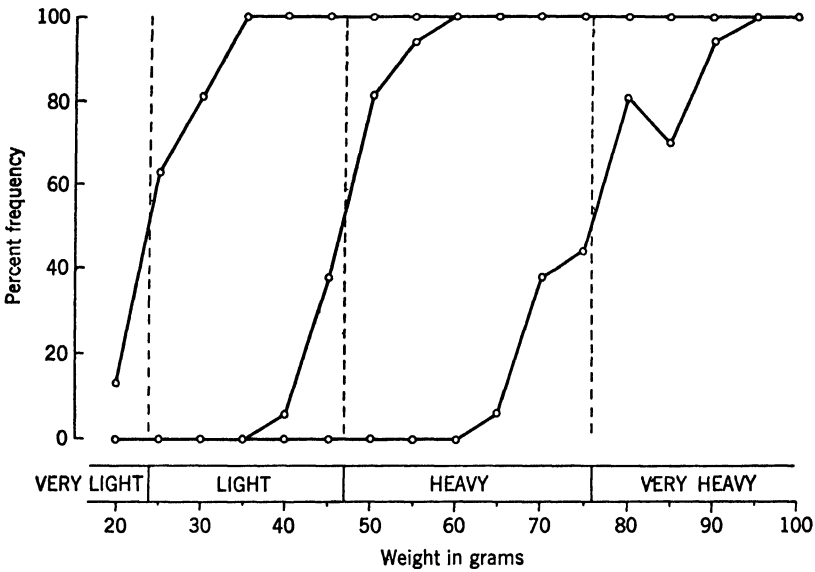


FIGURE 10.5. Transition zones and category limens for a four-category response scale. One subject judged each weight 16 times. From Johnson (*1944*).

This category limen is not a fixed point but the center of a transition zone which in this experiment extended from 30 grams to 90 grams. Within this transition zone responses fluctuated, a weight being called *light* sometimes and *heavy* at other times. The mean value of 48 grams for the limen is calculated from the distribution of judgments within the transition zone by the summation method (see Woodworth, *1938*). The standard deviation of the distribution of judgments was 10 grams. Figure 10.4 shows the obtained cumulative frequency curve.

Another subject judged the same weights in four categories: *very light, light, heavy, very heavy*. Thus we have three transition zones and can compute three limens: 24, 47, and 76 grams. Figure 10.5 shows

the cumulative frequency curves and the response scale computed from them.

A by-product of these computations is the quantitative definition of the response categories. The category *light,* for example, extends from 24 to 47 grams. The category *heavy* extends from 47 to 76 grams. As this subject made these judgments, then, he defined *light* as a category 23 grams wide and *heavy* as a category 29 grams wide.

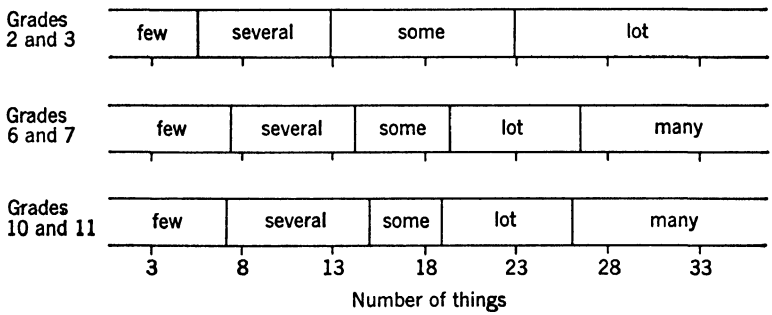


FIGURE 10.6. Quantitative description of response scales used by children to designate various numbers of printed characters. Data from Brotherton, Read, and Pratt (*1948*).

This technique of analysis may yield useful information when other techniques are not applicable. Brotherton, Read, and Pratt (*1948*), for example, wanted to know what school children mean when they use indefinite number concepts like *several* and *many.* They printed various numbers of typographical characters, e.g., 8 dollar signs, 28 asterisks, in rows and asked their subjects to check each as *hardly or scarcely any, few, several, some, lot,* or *many.* The data are somewhat irregular and hard to summarize by conventional procedures, but the response scale can be neatly represented by the procedure described above. The cumulative frequency distributions for the five transition zones are somewhat skewed, but rough limens can be easily obtained by plotting the data on probability paper. The response scales thus computed for three groups of subjects are shown in Figure 10.6. This procedure does not distort the data in any important way, but shows how many things the subjects actually referred to when they used this scale of indefinite number concepts.

Some very useful data for our purposes, now and later, have been published by Philip (*1947*) in the form of responses on an 11-point

scale. His summarized description of the experiment follows, in which O is a symbol for observer.

Seven O's made a prolonged series of judgments of colour mass on material which was tachistoscopically exposed for brief controlled periods. The material consisted of six sets of cards, each of which comprised eleven cards printed in two colours to form a graded series such that, as one colour gradually increased from card to card the other diminished. Each card held thirty-six evenly spaced and randomly distributed dots in either of two colours, blue and green. On one card, there were twenty-three blue dots and hence thirteen green dots; on the next card in the series, twenty-two blue dots and fourteen green dots; on the next card, twenty-one blue and fifteen green; the end card of the series had thirteen blue dots and twenty-three green dots. As each card was exposed, O had to determine the relative amount of a previously designated colour, and locate on a scale from one to eleven the point which, in his opinion, represented the relative amount of the designated colour. Hence, the psychometric procedure used was that of the absolute judgment. Preliminary practice was given before the experiment proper, and also before each session. A total of 1,650 judgments was obtained from each O, yielding some 11,550 judgments in all.

The number of times each stimulus card was judged in each category of the response scale is given in Table 10.1, taken directly from Philip and rearranged slightly. The stimulus cards are designated by the number of green dots; subtracting this number from 36 gives the number of blue dots (see Figure 10.7).

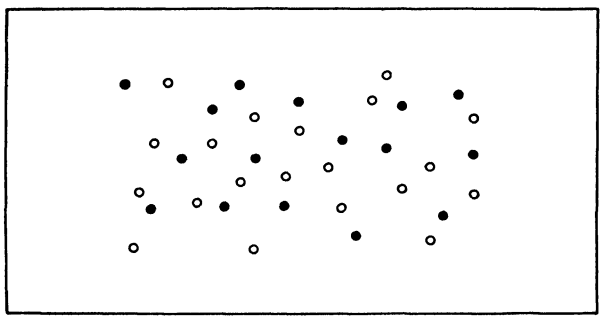


FIGURE 10.7. An illustration of the stimulus cards used by Philip (*1947*) for judgments of color mass. Green dots are shown as filled circles, blue dots as open circles. This is Card No. 16, since it has 16 green dots.

TABLE 10.1. Distribution of Judgments of Color Mass.[a]

Stimulus Number	Frequency of Judgments in Each Category										
	1	2	3	4	5	6	7	8	9	10	11
23	2	4	9	18	44	48	67	157	173	235	293
22	2	5	16	38	67	79	117	184	173	190	179
21	12	14	30	53	69	82	148	184	191	155	112
20	5	17	48	57	87	116	163	203	166	116	72
19	7	25	49	80	97	152	186	176	126	86	66
18	8	28	55	118	183	178	163	147	86	53	31
17	55	77	163	193	185	148	122	55	31	12	9
16	50	87	170	217	181	139	80	63	23	26	14
15	76	116	207	211	177	108	74	40	17	19	5
14	161	186	208	185	141	61	49	29	8	12	10
13	254	196	201	132	106	63	49	18	10	12	9
Totals	632	755	1156	1302	1337	1174	1218	1256	1004	916	800

[a] From Philip *(1947)*.

To extend the previous treatment of scales to this 11-category case it is first necessary to reduce these frequencies to proportions. Each card was judged 1050 times, and the card with 13 green dots was placed in the first category 254 times, which is 24.2 percent of all judgments of that card. Transforming all entries in Table 10.1 to proportions and cumulating from the right or "greater than" end, we get Table 10.2, showing the relative frequency with which each stimulus card was judged higher than each designated category. As an illustration, 94 percent of the judgments of No. 22 placed it in category 5 or higher.

TABLE 10.2. Cumulated Relative Frequencies Computed from Table 10.1.

Stimulus Number	Proportion of Judgments "Greater Than" Each Category									
	1	2	3	4	5	6	7	8	9	10
23	.996	.992	.983	.966	.924	.878	.814	.665	.501	.278
22	.996	.991	.976	.940	.876	.801	.690	.515	.351	.170
21	.986	.973	.944	.894	.828	.750	.609	.434	.253	.106
20	.993	.977	.931	.877	.794	.684	.529	.336	.178	.068
19	.993	.969	.922	.846	.754	.609	.432	.265	.145	.063
18	.990	.963	.911	.799	.625	.456	.301	.161	.079	.029
17	.946	.873	.718	.534	.358	.217	.101	.049	.020	.009
16	.952	.869	.707	.500	.328	.196	.120	.060	.038	.013
15	.927	.817	.620	.419	.250	.147	.077	.039	.023	.005
14	.847	.670	.472	.296	.162	.104	.057	.029	.021	.010
13	.759	.572	.381	.255	.154	.094	.047	.030	.020	.009

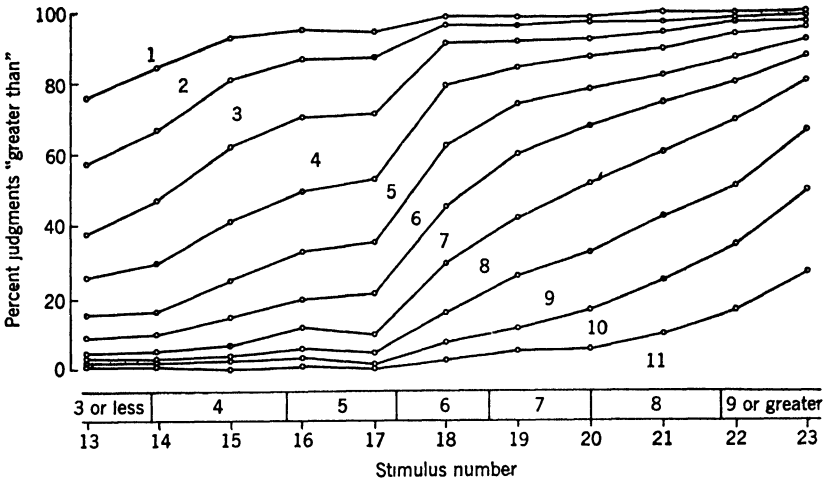


FIGURE 10.8. Cumulative frequency functions and category limens for an 11-category response scale. Group data from Philip (*1947*).

These ten cumulative frequency functions, graphing the transitions from each category to the adjacent category, are shown in Figure 10.8.

The category thresholds, which are the means of these ten psychometric frequency functions, can be computed by an extension of the usual techniques for treatment of psychophysical data obtained by the constant stimulus method. Summation methods are not applicable because a complete series of frequencies extending from 0 to 100 percent cannot be assumed. But normal ogives can easily be fitted to the data, and means and standard deviations can be computed in the conventional way. Plotting the data on probability paper shows that the distributions are approximately normal and yield means that are almost identical with those computed arithmetically.

From Figure 10.8 it is apparent that the three categories at each end of the response scale were not often used. The lowest category threshold, between categories 3 and 4, was 13.9. That is, at this point responses of 3 and lower were made with the same frequency as responses of 4 and higher. The threshold between categories 2 and 3 cannot be computed, except by a meaningless extrapolation, because responses of 2 and lower do not reach 50 percent. In the same way 21.8 is the stimulus value which was described as 8 or lower and 9 or higher equally often. The objective stimulus scale is 10 units in length,

extending from 13 to 23. The determinate portion of the response scale is 8 units in length, extending from 13.9 to 21.8.

When the response scale is described thus in stimulus units, the width of each category of the scale can be determined by simple subtraction. Categories 4 and 8 are 46 and 40 percent wider than the middle category, No. 6.

It is important to note that these calculations make no assumptions about the response scale. They simply describe how it was used. Though the categories of this scale used by Philip's subjects are numbered, these numbers are not employed in the calculations, and previous illustrations have used qualitative unnumbered categories. Rather, all calculations are in the units of the stimulus scale, and this procedure is applicable whenever the stimulus variable is suitable for quantitative treatment.

AVERAGE STIMULUS IN EACH CATEGORY. Some information about the response scale can also be obtained by tabulating the values of the stimuli assigned to each category and computing the averages of these. Figure 10.9 illustrates the application of this method to Philip's data. Since the distributions of stimuli in the end categories are usually skewed, the use of medians should also be considered.

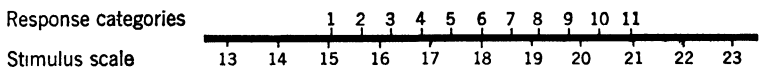


FIGURE 10.9. Locating response categories on the stimulus scale according to the mean stimulus placed in each category. Data from Philip (*1947*).

This method does not really describe the response scale, for it does not tell us what category will be used for any stimulus, nor the width of any category. Average stimulus value can always be calculated, however, for any category that is used at least once.

REPRODUCTION METHOD. The response scale can also be described by a reproduction or adjustment method. Almost everyone has a scale of heights of men and can hold his hand at a height above which he would call men *tall* and below which he would call them *short*. Or the subject may be asked, after listening to a series of sounds, to adjust the sound source so as to produce a sound at the center of the series, between *low* and *high*.

In one experiment (Johnson, *1946*) the subjects pitched pennies to

a wall and judged each throw *much worse than usual, worse than usual, better than usual,* or *much better than usual.* After some practice each subject marked his scale of values on the floor by placing a penny at each of the three boundaries between adjacent categories.

Data can be obtained by this method with as much detail and precision as desired. Kaufman, Reese, Volkmann, and Rogers (*1947*) asked their subjects to adjust a line to angular inclinations of 1, 3, 6, 10, etc., degrees from the vertical and calculated the means of these adjustments. From these data the subjects' response scale can be described,

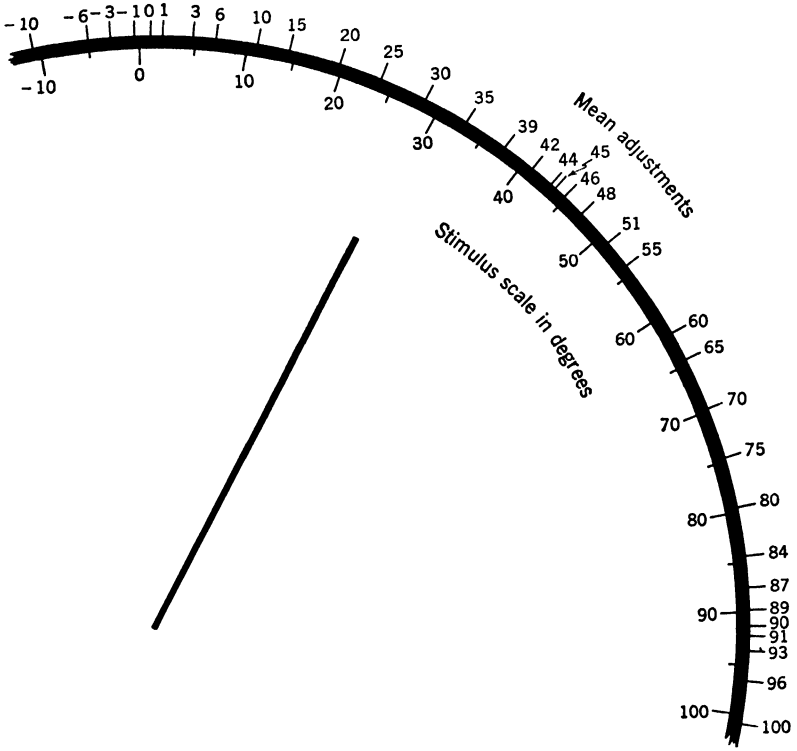


FIGURE 10.10. Relation between stimulus scale and response scale described by a reproduction method. The subjects were asked to adjust a line of light to an angle of, e.g., 30° from the vertical. Mean adjustment was 28.6°, as shown on outside scale. When 3° was called for, the mean adjustment was 4.5°. There were six subjects and 2870 judgments. Data from Kaufman, Reese, Volkmann, and Rogers (*1947*).

as in Figure 10.10, and the constant errors, variable errors, anchoring effects, and the like can be computed.

THE RATING METHOD. The most common way of expressing judgments is in terms of a rating scale, such as *very low, low, medium, high, very high,* or 1, 2, 3, 4, 5, 6, 7. If we assume that these category names represent equal units on the rating scale, the ratings of each stimulus object can be tabulated and statistically manipulated in the usual way. Philip (*1947*) used this rating method in his analysis of the data, previously described, on judgments of color mass. Statistical constants for the distributions of 1050 judgments of each stimulus

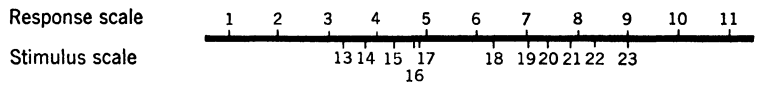


FIGURE 10.11. Locating stimulus objects on the rating scale according to the mean rating of each stimulus. Data from Philip (*1947*).

card, as computed by Philip, are shown in Table 10.3. Figure 10.11 shows the location of each stimulus object on the response scale according to the mean rating.

TABLE 10.3. Distribution of Ratings of 11 Stimulus Cards of 36 Green and Blue Dots, Numbered According to the Number of Green Dots [a]

Stimulus number	13	14	15	16	17	18	19	20	21	22	23
Median rating	2.9	3.4	4.1	4.5	4.7	6.3	7.1	7.7	8.1	8.6	9.5
Mean rating	3.3	3.7	4.3	4.7	4.8	6.3	7.0	7.4	7.8	8.3	9.0
Standard deviation	2.2	2.1	2.1	2.1	2.0	2.1	2.3	2.3	2.3	2.0	2.0

[a] Computed by Philip (*1947*) from the data of Table 10.1.

This method assumes that the categories of the response scale are equal. It makes no assumption about the stimulus objects and can be used when the stimulus objects are not suitable for quantitative treatment. Distributions of ratings of the end stimuli are usually skewed, hence the use of median ratings may be preferable to mean ratings.

EVALUATION OF METHODS. The merit of any method of computation depends, of course, on the purpose for which the results are to be used. Measurement of category limens is the method of choice if the purpose is to describe the scale that the judge actually uses and to locate the response categories on the stimulus scale. As we shall see later, some of the important phenomena of judgment can only be understood in relation to the category limens. This is the only method that yields Figure 10.8, from which we can read how large a stimulus must be to reach the threshold of any category.

Computing the average stimulus in each category does not give any information about category boundaries, but does show the typical stimulus placed in each category. This method may often be used when it is not possible to compute category limens for the end categories. This method and the measurement of category limens have both been used and the difference between them discussed by Taubman (*1950*) in a study of number judgments.

Computing the average rating of each stimulus is the reverse of computing the average stimulus placed in each category. The former assumes that the categories are equivalent, while the latter assumes that the stimuli are measurable. When stimuli measured in physical units are judged in categories of equal units, these two methods give the same results, and both are comparable to results obtained by the measurement of category limens.

The reproduction method has much to recommend it if adjustment of the stimuli is convenient. Interpretation is quite straightforward. Results obtained by this method seem to be similar to results obtained by the other methods (Kaufman, Reese, Volkmann, and Rogers, *1947;* Tresselt and Becker, *1950*), but the data have not been obtained under conditions that permit an exact comparison.

In the pages that follow all methods will be found useful for one purpose or another.

ORGANIZATION OF THE RESPONSE SCALE

The preparatory set, established usually by instructions, determines the stimulus aspect to be judged and the category names in which the judgments are to be expressed. The other important variable influencing the judgments is, of course, the stimulus variable. The next task, therefore, is to describe how the stimuli presented for judgment influence the response scale on which the judgments are made.

When the judge comes to the judgment with a very familiar response scale, the influence of the stimuli at hand on the response scale may be negligible. This is the case when one judges heights in feet and inches. The best information on the organization of a response scale comes, not from the use of a well-established scale, but from experiments in which a new scale is organized out of experience with a distribution of stimuli under the experimenter's control.

It is important to note that when the set restricts attention to specified stimulus objects, it fixes the context of judgment. If a hundred people are asked to judge Hitler's contribution to world history, the contexts within which this object is appraised will vary tremendously. But if these hundred people are presented a dozen simple homogeneous stimuli, such as lines, weights, or sounds, for judgment several times, the distribution of stimuli will usually operate as the context of judgment for all judges and each stimulus will be judged within this context. Thus the context can be experimentally controlled. Much of our knowledge of the formation of response scales comes from investigation of the relations between a distribution of homogeneous objects of judgment and a distribution of judgments of these.

When coöperative subjects are asked to rate the pleasantness of a series of color cards on a 7-point scale, or to judge a series of weights as 1, 2, or 3, or to rate color mass on an 11-point scale, they usually adjust their response scales to the distribution of stimuli quite readily. They make their judgments without difficulty, using the high categories of the scale for the high stimulus values and the low categories for the low stimulus values. When teachers grade students' papers on a scale of 1 to 10 or A, B, C, D, and E, they usually move their scales up or down to correspond to the range of achievement at hand. All the illustrations used in the preceding section on methods of describing the response scale, and data to be discussed below, show that when the stimulus material is unfamiliar and it is judged on a new response scale, there is at least a general correspondence between the two. This principle of general correspondence may be stated as a first principle of the organization of a response scale: *new scales based on experience with a homogeneous series of stimuli correspond roughly with the distribution of stimuli presented for judgment.*

Hunt and Volkmann (*1937*), discussing the results of an experiment on affective judgment, phrased this general principle as follows:

The position and the width of this scale are determined by the position of the group of stimuli and by the stimulus-range which the group covers. In general, when the group of stimuli is moved up or down upon the stimulus-continuum, the absolute scale moves with it; when the group of stimuli expands or contracts, the scale likewise expands or contracts. There is still variability of judgment, but the general position of the scale is determined.

This general correspondence between the response scale and the distribution of stimuli seems rather obvious, but it has great significance for psychological theory. Looking at this phenomenon from a broad biological point of view, we see once again the selective capacity of the organism. The judge is able, following instructions, to limit his perception to specified stimulus objects and his responses to specified category names. Within these limits we see the generalizing or conceptualizing capacity of the organism, for the responses are systematically adjusted to the stimuli. The correspondence is not usually between any particular response and any particular stimulus, but rather between a system—more precisely, a scale—of responses and a distribution of stimuli. We have long known that the organism can adjust to its environment, and that under some conditions it can learn to make specific responses to specific stimuli. Now we see that human beings under certain social conditions can acquire a high-level abstract adjustment very quickly. They can adjust a segment of the response repertoire to a segment of the environment in an orderly way with very little practice. This is a kind of adaptation to the environment that psychological theory must take seriously.

There have been two serious attempts to show how the response scale gets tied to the distribution of stimuli. Johnson's generalization theory (*1944, 1945, 1946, 1949a*) treats the formation of a scale as a higher form of learning or conceptualization. Helson's formulation (*1947, 1948*) is an extension of his theory of adaptation-level, which "originated as a shorthand description and explanation of certain fundamental phenomena in vision," to this problem area. These formulations, which despite their separate origins turn out to be similar, can be set forth most conveniently by considering first the centering of the response scale and second the spacing of the scale values of the response scale in relation to the stimulus scale.

CENTERING THE SCALE

The center of a two-category scale is, of course, the limen between the two categories. The center of a four-category scale would be the limen between the second and third categories. The center of a five-category scale would presumably be the middle of the middle category. The principles are the same for all scales, but the two-category scale can be discussed most simply.

Series of stimuli that vary along a single stimulus dimension, e.g., weights, sounds, are presented to a subject one by one, the method of single stimuli, with instructions to judge each in one of two categories, e.g., *low* and *high*. The subject will be confused at first by the ambiguity of the task and his judgments will be quite variable. Soon, however, the variability will be reduced and a tabulation of the judgments will permit the inference that he has developed and is using a two-category scale related in some consistent way to the distribution of stimuli. The theoretical task is to account for the center of the scale, the limen between the two categories, in terms of the distribution of stimuli in the series presented for judgment. The assumption is made that the scale is developed within the experiment and is not seriously influenced by previous experience.

The generalization theory (Johnson, *1944*) begins with the statement that the effect of any stimulus, x, on the perceiving organism is functionally related to that aspect or dimension of the stimulus which the organism is prepared to perceive. Calling the central effect of stimulation y, we can write an equation for the receptor function:

$$y = f(x) \qquad (1)$$

The stimuli are measured in x-values or physical units, but it is their central effects or y-values that are organized into a response scale. To illustrate how the organization occurs, let us assume that two stimuli, x_l and x_h, are presented to the subject with instructions to call each *low* or *high*. This is the simplest sort of pattern, but a pattern nonetheless, because x_l is low only in relation to x_h and x_h is high only in relation to x_l.

Then another stimulus, x_o, located between x_l and x_h, is presented to the subject with instructions to call it *low* or *high*. Since only two categories of response are permitted, the question is: Will x_o evoke the same response as x_l or as x_h? The generalization theory states that the effect of stimulation by x_l, which we call y_l, spreads upward and

downward on the perceived stimulus continuum, and that the effect of stimulation with x_h, which we call y_h, spreads likewise (see Figure 10.12). The tendency to equate or assimilate x_o to x_h and thus to call it *high* will be greater the closer y_o is to y_h; i.e., the farther it is from y_l. In the general case, when the subject has been exposed to many

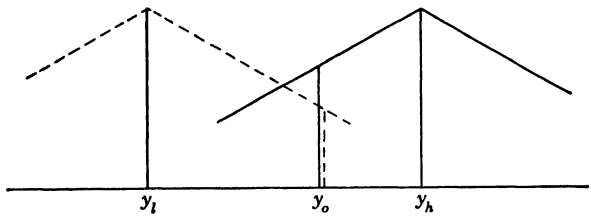


FIGURE 10.12. Showing generalization gradients assumed in Equations 2 and 3.

stimuli, x_i, yielding the central effects, y_i, if the generalization gradient is symmetrical about the point of origin, the tendency to call any particular stimulus, x_o, *high* is a function of the distance of the central effect of that stimulus, y_o, above the central effects of all stimuli, y_i.

$$T_h = F(y_o - y_i) \tag{2}$$

If we assume that the effect of stimulation by any stimulus spreads up and down the continuum in a linear way, F becomes a simple constant of proportionality and we can sum these tendencies for N stimuli.

$$\Sigma T_h = k(y_o - \Sigma y_i) \tag{3}$$

To locate the point on the stimulus continuum where the resultant tendency to respond *high* is equal to the tendency to respond *low*, i.e., the category limen, we set this quantity at zero and solve for y_o. When $\Sigma T_h = 0$,

$$y_o = \frac{\Sigma y_i}{N} \tag{4}$$

Equation 4 states that the midpoint of a two-category response scale, the point at which a stimulus will be judged in one category as often as the other, is the simple arithmetic mean of the central effects of the stimuli upon which the scale was constructed.

But these central effects or y-values are not measurable. They are

related to the measurable x-values by the receptor function, Equation 1. The receptor function for judgments of lifted weights is apparently quite complicated, but for judgments of pitch the function is usually assumed to be a simple logarithmic one: $y = f(\log x)$. Hence it follows that for judgments of pitch the midpoint of the response scale, which we may now call L for category limen, is the logarithmic or geometric mean of the stimuli presented for judgment.

$$\log L = \frac{\Sigma \log x_i}{N} \tag{5}$$

Thus we return to the stimulus objects, measured in stimulus units. Equation 5 is in a form that can be verified by experiment. The limens can be computed by the procedure described earlier in this chapter, and the stimuli are measured in physical units, e.g., grams or cycles per second. Figure 10.13 shows that, when a wide variety of distributions of sounds were judged as to pitch on a two-category scale, the obtained category limens agreed quite well with the predicted limens. It is worthy of note that in this case of judgments of pitch where a simple logarithmic relation is assumed, the theory permits genuine prediction and a rigid test of the outcome. There are no empirical constants to be calculated; the predicted limens can be computed before a single subject enters the laboratory. We are not fitting a curve but computing a single value for each series of stimuli, to which a single obtained value can be compared.

Computations of this kind make several assumptions, and in reverse these assumptions are supported to the extent of the accuracy of the predictions. As already mentioned, we assume that the scale is organized during the experiment and is independent of previous experience. In most experiments this assumption seems to hold quite well; other cases will be discussed below.

Another assumption, implicit in Equation 4, is that each judgment of a stimulus can be treated as a unit of stimulation or practice. When some stimuli are presented oftener than others, or emphasized more than others, appropriate weighting will be necessary.

A third assumption, central to the theory, is that the category limen is not a response to any one stimulus but is a generalization based on all stimuli judged. The center stimulus, if there is one, has no particular significance in the calculations. We assume further that this generalization takes place along linear gradients, as in Equation 3

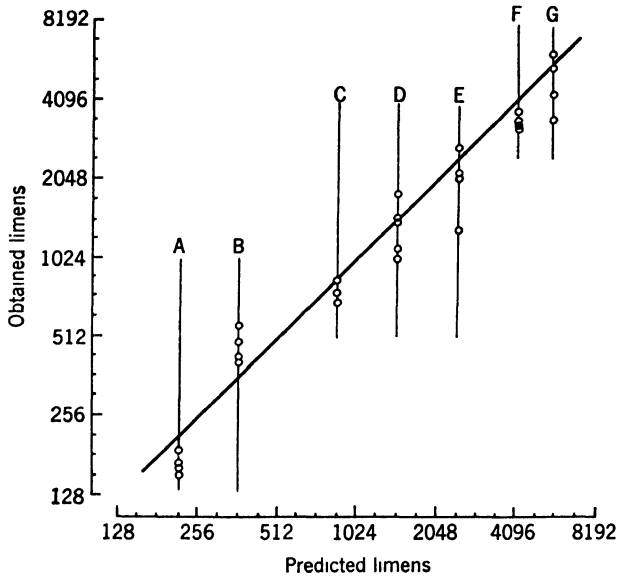


FIGURE 10.13. Comparison of predicted and obtained category limens for two-category judgments of pitch. The vertical lines represent the range of frequencies used in a series. The 29 points represent the obtained category limens or boundaries, stimuli above which were judged *high* and below which were judged *low*. The obtained values are plotted above the values predicted by Equation 5 so that if theory and fact agreed perfectly, all points would lie on the diagonal. The correlation between the two is .97. From Johnson (*1949*).

and Figure 10.12. To test this assumption of linearity of spread, skewed distributions of stimuli were used in experiments on judgments of weights (Johnson, *1944*) and judgments of pitch (Johnson, *1949a*). If the generalization gradients were not approximately linear, the predicted limens would have differed systematically from the obtained limens. Hence the results in both experiments support the assumption of linearity of spread.

Equation 4 represents a general relationship. Equation 1 is specific to the stimulus dimension: for judgments of pitch it is logarithmic. The buffer action of receptor mechanisms suggests a logarithmic relationship, or some similar decreasing function, for many perceived

dimensions, such as brightness, loudness, and weight. For judgments of lengths of lines, number of dots, and the like a linear relationship can be assumed. For Philip's data on judgment of number of colored dots a linear relationship is most reasonable. Therefore the center of the scale would theoretically be the arithmetic mean of the stimulus series, which is 18. The middle category of the response scale extends from 17.3 to 18.6. Hence the center is actually 17.95, and theory and fact agree very well. Use of the geometric mean in this case would lead to a predicted center of 17.7.

An important consequence of the shape of the receptor function, according to the generalization theory, is that when it is a decreasing function, the center of the response scale will be lower than the center of the stimulus series. When it is a linear function, the two centers will coincide. If the receptor function were an increasing function (no such receptor function is known, but pain is a possibility), the center of the response scale would be higher than the center of the stimulus series. The facts, as far as they are known, support this deduction.

Helson's formulation of these facts (*1947*) began with visual phenomena, hence the term adaptation-level. As applied to two-category judgments, adaptation-level and category limen are operationally identical. The adaptation-level or *AL* is assumed to be a function of all stimuli acting on the organism and, in accordance with the Weber-Fechner law, is calculated as a geometric mean. For color vision data the background would be weighted three times as much as the stimulus, but for judgments of series of homogeneous stimulus objects, such as lifted weights, the stimuli are weighted equally. The assumption of linearity of spread, mentioned above, is not explicitly made, but the averaging process would seem to imply such an assumption and the use of skewed distributions of stimuli as a test of the theory is welcomed.

Since *AL* from some distributions of stimuli is known to be below the geometric mean, Helson adds an empirical constant, *d*, said to be related to time-order effects and the size of the step-interval, to the left side of the equation. Thus Helson writes

$$\log (AL + .75d) = \frac{\Sigma \log x_i}{N} \tag{6}$$

When there is no time-order error, as in judgments of pitch (Postman, *1946*), and when differences in step-intervals may be assumed

to be neglible, the empirical constant may be eliminated. Equation 6 then becomes identical with Equation 5.

These two quantitative formulations of the centering of a response scale are similar in being based on an averaging process. In the case of lifted weights, both use a kind of geometric mean. But the geometric means are higher than the obtained category limens or adaptation-levels. The generalization theory corrects for this discrepancy by assuming an empirically modified geometric receptor function, while the adaptation-level theory subtracts an empirical constant from the *AL*. Using Equation 6, which includes the empirical constant, Helson (*1947*) is able to predict Johnson's lifted-weight data of 1944 quite well. For judgments of pitch the generalization theory is adequate (Johnson, *1949a*). The adaptation-level theory would be equally adequate if the constant, *d*, of Equation 6 is assumed to be zero.

The generalization theory has also been applied to the scale that one organizes for evaluating success and failure (Johnson, *1946*). Each of ten college women pitched a penny to a wall 50 times and judged each throw *much worse than usual, worse than usual, better than usual,* or *much better than usual.* In this experiment, therefore, the object of judgment, the distance of the penny from the wall, was not under the experimenter's control but was the result of the subject's performance, and each subject's scale was different. Each subject periodically indicated her three limens between the four categories by placing a penny on the floor where it would separate throws *worse than usual* from throws *better than usual,* and so on. The center of each subject's response scale was predicted from records of the throws by Equation 4, assuming a linear receptor function, and compared with the values obtained directly from each subject by the reproduction method. Although the throws varied over a range of 30 to 40 inches, the average error in predicting the center of the scale was only 2.1 inches for these ten subjects. The constant error of prediction was 0.8 inch. In this case any procedure that included the use of a geometric mean would yield large errors of prediction.

Shifting the Scale. In all of these experiments the same stimulus objects have been presented repeatedly, sufficient practice being given that the relation of the response scale to the distribution of stimuli becomes stabilized. Now suppose the stimulus series is shifted upward. Will the response scale shift also? And if so, how rapidly? Suppose, for example, we ask a friend what is the average price of a good lunch

today and, on the basis of his contemporary experience, he says, "One dollar." If inflation sends prices up, will his estimate of the average price of a good lunch rise with the times, or remain tied to his previous experience? Or, one might ask, what will happen to a college student's scale of the seriousness of crimes if he loses contact with the peccadilloes of undergraduate life and goes to work in a state prison?

Tresselt (*1947*) attacked this question by having her subjects judge one series of lifted weights, then a lower series. Different groups of subjects had different amounts of practice with the first series before shifting to the second. Basing her conclusion on the mean values of the weight judged *medium,* she states that "the larger the period of practice the more slowly does the scale of judgment shift to its new position."

Johnson (*1949a*) has attempted to extend the generalization theory to this case in which the stimulus series is shifted. For judgments of the pitch of sounds Equation 5 locates the limen of the two-category scale. It is simply the geometric mean of the frequencies presented for judgment. If higher frequencies are presented, the limen would be expected to rise. If we present the subjects with two series of sounds, the predicted limens of which are L_1 and L_2, we may expect that the limen will shift from L_1 toward L_2. The simplest expression describing the shift would be a weighted average of the two limens.

$$L_n = \frac{wL_1 + nL_2}{w + n} \tag{7}$$

In this equation n represents the number of times the second series of stimuli has been presented for judgment. L_2 is weighted by n, therefore, and L_1 is weighted by w, an empirical constant indicating the weight of the first or preshift series in impeding adaptation to the second or postshift series. L_n is the category limen or adaptation-level at the time of the n^{th} practice trial, i.e., the n^{th} presentation of the second series. The expression could be considered a moving weighted average. It begins at L_1 and takes the form of a hyperbola with L_2 as an asymptote.

When experiments were arranged to test this equation, the results of Figures 10.14 and 10.15 were obtained. They show that a hyperbolic function is not inconsistent with the facts, but the data are not sufficiently regular to rule out other similar functions that approach

an asymptote. More important, these graphs show that w, the empirical constant indicating the weight of the first series in impeding adapta-

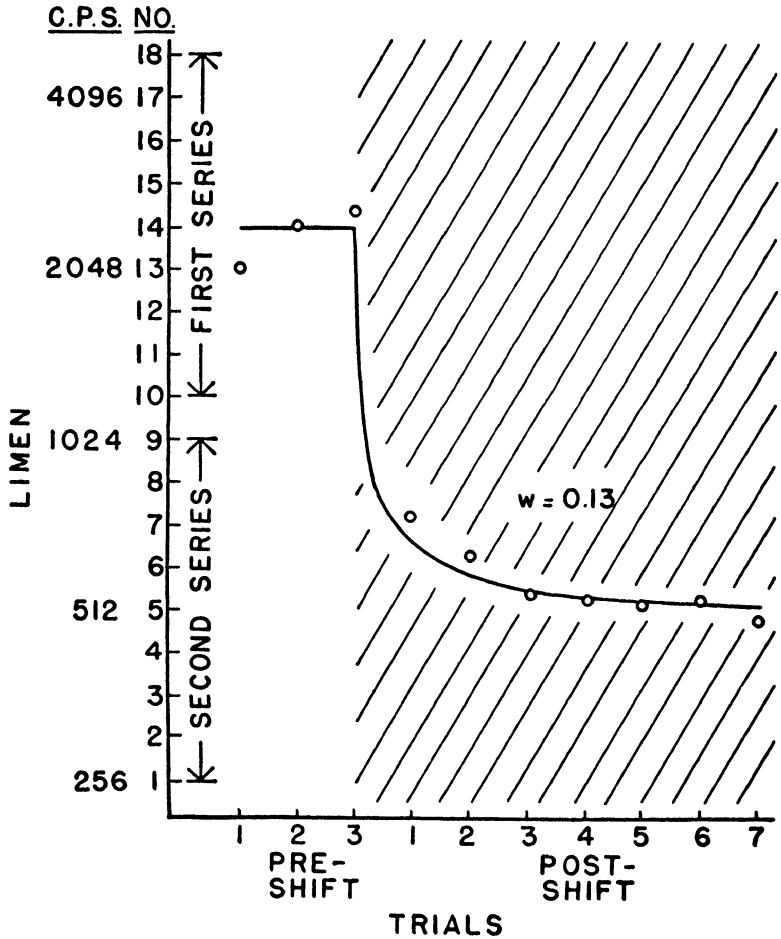


FIGURE 10.14. Learning curve for a downward shift in a scale used for judging pitch. The ordinate may be read in cycles per second or in quarter octaves. The circles represent limens of a two-category response scale: *low* and *high*. The theory locates the limen at 14 for the three trials with the first series, then it rapidly approaches 5 as an asymptote in the seven trials with the second series. The agreement of theory and fact may be judged by the deviations of the circles from the curve. From Johnson (*1949a*).

tion to the second, increases with an increase in the number of pre-shift trials.

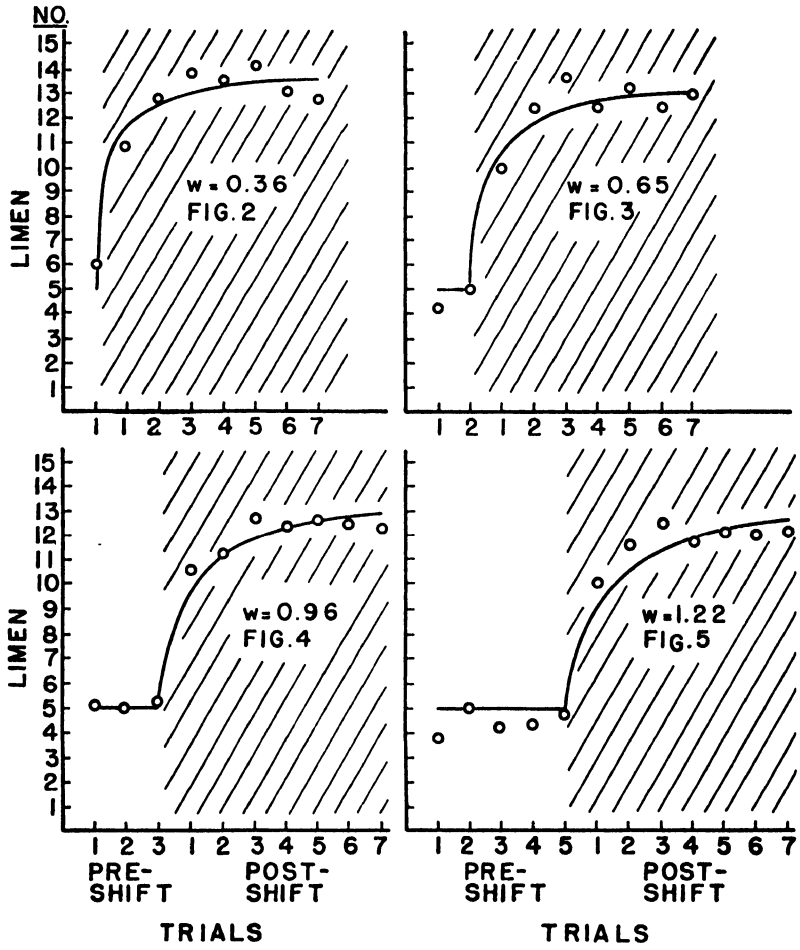


FIGURE 10.15. The effect of different amounts of practice with the first series on rate of shifting to the second series. The first series was presented 1, 2, 3, and 5 times and the second series was presented 7 times. The theoretical curve of Equation 7 locates the limen for the first series at 5, and with the introduction of the second series it approaches 14 as an asymptote. The weight of the first series in retarding adaptation to the second is indicated by the empirical constant w. From Johnson (*1949a*).

To anyone who knows how curves can be fit by a judicious selection of constants, it is worthy of note that all these data are fit by the same equation, with only one constant, w, to be empirically determined. The values of the limens L_1 and L_2 are computed by Equation 5, which includes no empirical constants. This case of a shift in the stimulus series could also be handled by a suitable modification of the adaptation-level theory.

These experiments by Tresselt and Johnson are based on the assumption that the response scales are new scales, organized out of laboratory experience with the stimulus series. To check on the influence of previous extra-laboratory experience on the organization of scales in the laboratory, Tresselt (*1948*) persuaded some student watchmakers, used to lifting small weights, and some amateur weight lifters, used to lifting heavy weights, to make judgments of laboratory weights. As expected, the watchmakers' scales were lower than the weight lifters' scales. The scales for these two groups did not come noticeably closer as a result of laboratory practice, so the difference may not be entirely a result of practice.

Centering Around a Standard. When comparative judgments are made by the method of constant stimuli, each of a series of stimuli is compared with a standard stimulus. Judgments of lifted weights, for example, are expressed as *lighter* or *heavier* than the standard. It has long been known, however, that the center of this two-category scale does not exactly coincide with the standard stimulus. It is usually lower. If weights of 92, 96, 100, 104, and 108 grams are compared with a standard of 100 grams, the limen between the two categories will usually be around 98 or 99 grams. Much theorizing and experimentation have been directed toward an explanation of this constant error (see Woodworth, *1938*).

If we treat the comparative judgment as judgment on a two-category scale and apply the principles described above to the centering of the scale, instructive results are obtained. According to the generalization theory, the center of the scale is a function of all stimuli, including the standard stimulus. If the receptor function is a logarithmic function, or some similar decreasing function, the category limen will necessarily be lower than the standard—simply because the standard is located at the arithmetic mean of the series and the geometric mean is necessarily lower than the arithmetic mean. Just how much lower it will be depends on the exact shape of the receptor function and the relative

weight to be given to the standard stimulus. If we assume that the subjects judging the variable stimuli pay relatively less attention to the standard stimulus, we can give the standard a weight of half as much as the other stimuli in computing the theoretical limen. Computations of this sort are consistent with published data (Johnson, *1945*). Computations according to the theory of adaptation-level, weighting the standard stimulus one-third as much as the other stimuli, also agree fairly well with a wide range of data (Helson, *1947*).

It would be difficult to check these predictions about the comparative judgment with any precision because the relative weights of standard and variable stimuli can only be conjectured. Furthermore, the effects of stimulation often fluctuate between the presentation of the standard and the variable stimuli in a way that complicates the comparative judgment. But the direction of the constant error is clearly in agreement with predictions from any theory that involves a geometric mean. Comparative judgments of lengths of lines, for which a linear relationship may be assumed, do not show consistent constant errors.

LOCATING THE SCALE VALUES

Assuming that we have accounted for the coincidence of the center of the response scale and the center of some function of the stimulus series, the next task is to account for the relationship between all values on the response scale and all values of the stimulus series. The generalization theory and the adaptation-level theory and any other reasonable theory must assume that the relationship is not between specific stimuli and specific responses but between all stimuli and all responses. Any adequate theory must be based on some sort of generalization of, or adaptation to, the whole series of stimuli that are judged.

Some response scales, to be sure, are irregular and cannot be systematically related to the stimulus series. Some stimuli stand out from the others and some response categories are preferred over the others. But when the categories are approximately homogeneous, as when they are designated by numbers, theoretical analysis is profitable.

The most important general trend in the relation between the response scale and the stimulus values was discovered many years ago by Hollingworth (*1909, 1910*) in an investigation of the inaccuracy of movement and named the "central tendency." His subjects had the task of reproducing distances from a starting point by moving an

indicator along a slide. It turned out that in any series of reproductions the large distances were underestimated and the small distances were overestimated. In each series it was possible to locate roughly an "indifference point" near the center of the series at which there was no constant error. The term central tendency refers to the fact that estimates of stimuli at the extremes are shifted toward the indifference point at the center. Similar constant errors have been noted by Ipsen (*1926*) for judgments of lines, by Woodrow (*1933*) for judgments of weights, and by Turchoie (*1948*) for reproductions of short temporal intervals. When there is no quantitative stimulus variable, as in judgments of personality traits, we cannot speak of constant errors, but the tendency of most judges to use the middle categories of the rating scale more frequently than the end categories has been called the central tendency (Guilford, *1936*).

The facts may be clearly illustrated by returning to Philip's data on judgments of dots, described above. When an 11-point scale of 1 to 11 is used for repeated judgments of 11 evenly spaced stimulus objects, 13 to 23 dots, the correspondence between the response scale and the series of stimuli could be perfect. Stimulus 13 could be judged 1, Stimulus 14 could be judged 2, and so on to Stimulus 23, which would be judged 11. But Table 10.4 shows that Stimulus 13 received an average rating of 3.3, rather than 1. Stimulus 14 was rated 3.7 rather than 2. And the same regression toward the center of the scale occurs at the high end.

When category limens are computed from these data, central tendency shows in the extension of the limens outward from the center of the response scale and the vanishing of the end categories (see Figure 10.8). The limen between Categories 3 and 4 is not at a stimulus value of 15.5 but at 13.9. The limen between Categories 2 and 3 is shoved off the scale. And the same effect can be seen at the high end. Central tendency extends and truncates the response scale at both ends. These are the facts of central tendency that any theory of the organization of response scales must attempt to explain.

If we examine a large table of judgments, such as Philip has made available, instead of the smaller ranges of stimuli and the smaller scales of judgment used in most experiments, it is apparent that *this phenomenon of central tendency is another manifestation of the regression effect that always occurs when two series of data are not perfectly correlated.* A considerable simplification of some of the puzzling phe-

nomena of judgment is achieved when we treat this as a problem in correlation and regression.

TABLE 10.4. Theoretical and Obtained Data Testing the Regression Theory of the Central Tendency of Judgment[a]

Stimulus number	13	14	15	16	17	18	19	20	21	22	23
Theoretical rating	3.1	3.7	4.2	4.9	5.5	6.1	6.7	7.3	7.9	8.4	9.0
Obtained rating	3.3	3.7	4.3	4.7	4.8	6.3	7.0	7.4	7.8	8.3	9.0
Response category number	3		4		5		6		7	8	9
Theoretical category limen		13.6		15.3		17.0		18.7	20.4		22.1
Obtained category limen		13.9		15.8		17.3		18.6	20.0		21.8

[a] Calculated from preceding tables.

Let us begin by correlating the response values, which we shall designate as Y, with the stimulus values, which we shall designate as X. There are 11,550 pairs of values in Table 10.1, and the correlation between them, calculated in the conventional way, is .654. The regression equation for estimating Y, given X, is:

$$Y' = .590\,(X - 18) + 6.07 \qquad (8)$$

This regression line is graphed in Figure 10.16. The category limens, taken from Table 10.4, fall along this regression line, with small discrepancies. The mean ratings of each stimulus, taken from Table 10.4, likewise fall along this line (not shown), with small discrepancies. The explanation for the central tendency is now obvious. If the correlation between stimulus and response were perfect, as represented by the regression line for $r = 1.00$, there would be no central tendency. The central tendency or regression effect is measured by the discrepancy between the obtained regression line for $r = .654$ and the regression line for $r = 1.00$.

The importance of the correlation for central tendency can be appreciated by writing the regression equation in the form of z-scores. As

usual, z is the deviation of any value in a series from the mean of the series, and $\bar{z}_y$ is the estimate of the deviation of Y from the mean of the Y series when X is given.

$$\frac{\bar{z}_y}{z_x} = r_{yx} \tag{9}$$

Equation 9 shows that if $r = 1.00$, the judgment of any stimulus would be as far above or below the mean judgment as the stimulus is above or below the mean stimulus. As r decreases, the judgment of any stimulus moves closer to the mean judgment. The central tendency is $z_x - \bar{z}_y$

The final test of this regression theory of central tendency is a comparison of values computed by the regression equation with the obtained category limens and the obtained ratings (see Table 10.4). The theoretical ratings were obtained from the regression equation, entering the 11 stimulus values 13 to 23. The obtained ratings are copied from Philip's report, and are also shown in Table 10.3. The theoretical category limens are those values of the stimulus variable, X, which must be entered in the regression equation to get judgment scale values, Y', of 3.5, 4.5, 5.5 . . . 8.5. The obtained category limens were computed from Table 10.2. If we use the regression equation to get theoretical category limens for the end categories, we get impossible values, beyond the stimulus series. Thus the regression equation accounts for the vanishing of the end categories. Figure 10.16 and Table 10.4 show that the response scale values derived from the regression equation and the actual values of category limens and mean ratings are in very good agreement. The mean squared discrepancy between theoretical and obtained mean ratings is only .065.

Mean stimulus value for each judgment category can be predicted by the other regression equation, yielding 11 values of X for the 11 values of Y.

It is well known that any table of regular data can be reproduced by curve-fitting techniques if enough empirical constants are used. In the present formulation the data are fit by Equation 8, which contains only two empirical constants. The constant for the center of the scale, the intercept constant, can be computed theoretically, in advance of experimentation, if the conditions necessary for applying the generalization theory are met. In the section on centering the scale (see page 343), the center of this response scale from Philip's data was, in fact,

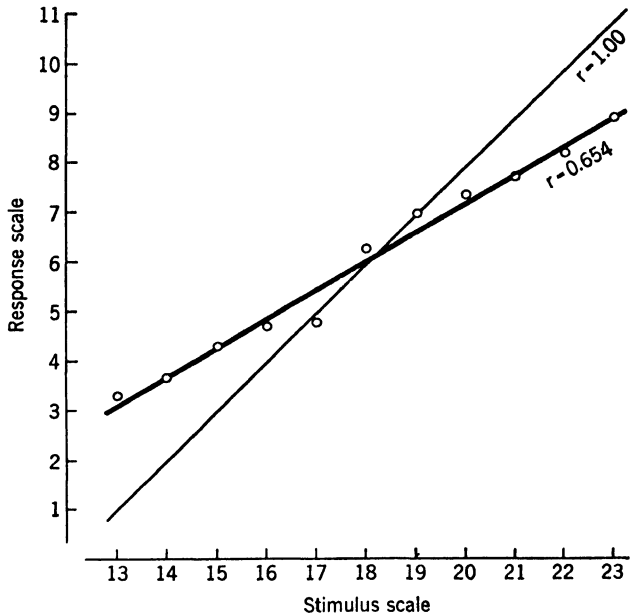


FIGURE 10.16. The effect of regression toward the mean on the response scale. The points are mean ratings taken from Table 10.3. The heavy line is the regression line for $r = .654$, which may be compared with the regression line for $r = 1.00$. Rotating the graph 90° will show how the lack of perfect correlation between response series and stimulus series pulls judgments toward the middle of the response scale.

theoretically computed by Equation 4 and a linear receptor function with very small error. Hence, under conditions in which Equation 4 holds, the regression theory of the relation between the response scale and the stimulus series requires only one empirical constant. This constant, the slope constant, is the product of the correlation coefficient and the ratio of the two standard deviations. It represents the accuracy of the judgments or, in other words, the general correspondence between the response scale and the distribution of stimuli. It depends on the variability of the judgments and is therefore influenced by individual differences. It must, of course, be obtained empirically.

Data on judgments of a short range of stimuli with-a short response

scale are not always suitable for computation of a correlation coefficient in the usual way. In such cases we can use the relationship

$$r^2 = 1 - \frac{\sigma^2_{y \cdot x}}{\sigma^2_y} \tag{10}$$

in which $\sigma_{y \cdot x}$ is the standard deviation of the judgments or ratings of a stimulus x, and σ_y is the standard deviation of all ratings. Combining equations 9 and 10, we can write

$$\frac{\bar{z}^2_y}{z^2_x} = 1 - \frac{\sigma^2_{y \cdot x}}{\sigma^2_y} \tag{11}$$

Since $z_x - \bar{z}_y$ is the regression or central tendency. Equation 11 shows that this effect is large when the variance of the judgments of any one stimulus is large relative to the variance of all judgments. If the data permit assumptions of linearity and homoscedasticity, Equation 11 can be used to compute the central tendency for series of three or more stimuli judged on scales of three or more categories.

A different manifestation of central tendency or regression can be seen in the data of Brotherton, Read, and Pratt, described on page 333. Various numbers of objects were judged by school children on a six-category scale: *hardly or scarcely any, few, several, some, lot, many.* The categories of the response scale are not of equal width, so they cannot be assigned numbers and correlated with the stimulus series. It is possible, however, to compute chi-square and a coefficient of contingency. This coefficient is .74 for Grades 6 and 7 and .73 for Grades 10 and 11. From a relation of this magnitude we would expect a little regression. That is, the regression theory would predict that the response scale would be extended and truncated. From Figure 10.6 we see that this did happen. The response scales for these grades have five categories instead of six. In Grades 2 and 3 the contingency coefficient was only .53. Hence in these grades we would expect more truncation, and in fact the scale for these grades has only four categories.

According to the above discussion, the central tendency could just as well be called a regression effect. When the response scale and the stimulus series are not perfectly correlated, the stimulus ratings regress toward the center of the rating scale and the category limens extend outward from the center of the stimulus scale. The magnitude of the regression is inversely related to the correlation between the two scales.

Or, in other words, the regression is large when the variance of the judgments of each stimulus is large. If the judge has to judge lines of 8, 9, and 10 inches as numbers 1, 2, and 3, the task is quite easy. There will be little variation in the judgments of each line and little central tendency or regression. The limen between Categories 1 and 2 will be about 8.5 inches, and that between Categories 2 and 3 will be about 9.5 inches. The 8-inch line will usually be called 1, the 9-inch line 2, and the 10-inch line 3. If the lines are more similar, the task is harder, and the judgments will vary. The middle category, 2, will be wider. Mean rating of the 8-inch line will be more than 8, and mean rating of the 10-inch line will be less than 10. In general the judgments are variable and regression effects appear when the stimulus objects are not clearly perceived or understood, when the response categories are not well defined, or when the two scales are not easily brought into correspondence.

Helson's theory of adaptation-level, which has been applied to the centering of a response scale, can also be applied to the location of points along the scale. Helson (*1948*) numbers the categories of the response scale, starting from the low end, and it is with these numbers, called P units, that the computations are carried out and the response scale (or rating scale) related to the stimulus series. The numbers representing repeated ratings of each stimulus are averaged to get a mean P for each stimulus. This averaging process assumes that the P units are equal or that inequalities are random, and Helson restricts his formulation to scales for which such an assumption can be made.

The theory states that the scale is organized around the adaptation-level, now called A rather than AL, and that any stimulus, x_i, will be located on the response scale according to its distance from A in units of just-noticeable-difference. According to Weber's law, the increment ΔX that must be added to X to obtain a just-noticeable-difference (j.n.d.) is a constant fraction of X, or

$$\Delta X = kX \tag{12}$$

where k is the Weber constant. But the scale is organized around A, the adaptation-level, so we are interested in ΔA. When a stimulus is judged in relation to A, we may say that two stimulus values are influencing the organism, X_i and A. The j.n.d. or ΔA, then, is a function of the average of the two.

$$\Delta A = k\frac{X_i + A}{2} \tag{13}$$

If we want to know how far any stimulus is from A in j.n.d.'s or P units, we can divide this distance in stimulus units by the j.n.d. unit; replacing $k/2$ with K, we have:

$$P = K\frac{X_i - A}{X_i + A} \tag{14}$$

The judge cannot rate X_i in j.n.d. units from A, but Equation 14 provides a way of using the customary ratings. There will be a mean rating or P for each stimulus, X_i, and from a series of these the center of the scale, A, and the other constants can be computed by the least-squares technique.[1]

The adaptation-level theory of the relation between the response scale and the series of stimuli and the regression theory described above are similar in that both are organized around a center and both use a least-squares technique to obtain certain constants. There are two main differences. (1) The AL theory, based on Weber's law, always assumes a geometric relationship between stimulus values and response values. The regression theory allows for differences in the stimulus material. The response-stimulus function is probably geometric for weight, brightness, loudness, and pitch, and linear for length and number. (2) The AL theory computes scale values by summing j.n.d.'s. The regression theory computes scale values by an ordinary regression equation, the constants of which depend on the variability of the judgments.

Applied to lifted-weight data judged on a nine-point scale, the adaptation-level theory fits very well (Helson, *1948*). After the two constants were calculated, the five theoretical P values agreed exactly with the five obtained mean ratings. Matching five points with the aid of two empirical constants is not conclusive proof; but as far as these results go, they support the theory for lifted-weight data, especially the assumption that the P units are approximately equal.

For a more exacting test Helson (*1948*) applied the AL theory to Philip's data, described above, on judgment of number of green dots in a mass of green and blue dots. In contrast to the linear relation shown in Figure 10.16, the AL theory results in a curvilinear relation.

[1] For further details of Helson's theory of adaptation-level, see Michels and Helson *(1949)* and Nash *(1950)*.

The ratings predicted by the curve of best fit are thrown off somewhat by the curvilinearity. The mean squared error of estimation is .168, as compared with .065 for the straight line of best fit. We must conclude that the simple assumption of linear correlation and regression accounts for the response scale used by Philip's subjects better than the complex theory of adaptation-level based on Weber's law.

If the generalization theory for centering the response scale is correct, the shape of the receptor function relating the stimulus variable to its effect on the organism is of primary importance. It is at this point that a nonlinear, e.g., a logarithmic, relationship may operate. Thereafter, if the regression theory for locating the scale values is correct, only simple assumptions and simple computations are necessary. If we picture the judge's task as one of fitting a series of response categories to a series of stimuli, Weber's law is irrelevant.

When the response scale and the stimulus series are fairly regular, regression accounts for most of the observed correspondence between the two. Irregularities in either variable produce some additional phenomena of judgment that can be described under two headings: response preferences and anchoring effects.

RESPONSE PREFERENCES

It has long been known that people prefer to make judgments in some categories rather than others. Census reports show a preference for multiples of five and for even numbers in reports of age. When sentencing criminals, judges prefer round numbers, as do psychologists and other scientists when reading laboratory instruments. The effect of a preference for any category is to increase the width of that category. In reporting age to the nearest birthday, for example, if the response scale were uniform, the category 49 would be used for ages between 48.5 and 49.5, 50 for ages between 49.5 and 50.5, and so on, each response category being one year wide. But if there is a preference for 50, this category will be used by some people of age 49 and some of age 51. Hence the category 50 might extend from 49.2 to 50.8, thereby attaining a width of 1.6 years, while the 49 and 51 categories shrink correspondingly.

Response preferences are particularly likely to appear when the categories are designated by qualitative terms, such as *very, some, and medium,* instead of numbers. Some judges seem to prefer extreme categories, and some seem to prefer the middle of the scale. But we

must remember that what has often been called a tendency to prefer the center of a scale is really the effect of regression when the response scale is not perfectly correlated with the stimulus series.

THE MIDDLE CATEGORY. Good evidence that a response preference alters the width of the response category comes from experimental manipulation of that preference. The evidence was collected in an attack on an old question of psychophysical methodology. In comparative judgments by the method of constant stimuli should the instructions permit the subject to use a middle category, such as *equal* or *doubtful,* or should he be restricted to two categories, such as *shorter* and *longer?*

Early discussion of this question dealt with the validity of introspective reports of the impression of equality. When this criterion was put aside and the *equal* response treated as the middle category of a three-category scale, the question became amenable to quantitative investigation. In a weight-lifting experiment Fernberger (*1930*) showed that when the instructions emphasized finding a difference between the standard and the variable weights, the width of the middle category was reduced to two grams; but when the instructions permitted free use of the *equal* response, the middle category attained a width of ten grams. The methodological implications of this discovery have been fully discussed by Woodworth (*1938*). Other conditions that affect the width of the middle category have been described by Cartwright (*1941a*). Some effects of scale numbering on scale reading have been described by Barber and Garner (*1951*).

ANCHORING

Under certain conditions, as noted above, the response scale is related in a generalized way to all the stimuli being judged. But certain response categories may be tied in a special way to certain stimuli. The lowest stimulus in a series, for example, may be easily identified as such. If so, it will always be placed in the lowest category of the response scale. The scale may then be described as anchored at the low end. This end is tied fast, while the remainder of the scale is tied more loosely to the remainder of the stimulus series. The principal anchoring effects have been described by Volkmann (*1951*).[2]

[2] See also a report (Mt. Holyoke Psychophysical Research Unit, *1953*) which summarizes work done between 1946 and 1952 for the Office of Naval Research on judgments of inclination, length, area, number, position, and shape. It contains valuable details on anchoring and discontinuities in scales of judgment.

When a certain stimulus ties down a certain category of the response scale because of some intrinsic character, it is called a *natural anchor.* A zero stimulus is often a natural anchor, as in some experiments reported by Blumenfeld (*1931*) on judgments of the irregularity of scattered dots on a card. There is no apparent limit to the high end of such a stimulus series, but at the low end a limit would obviously be reached when there is no scattering, i.e., when the dots lie in a straight line. The judgments indicated that the subjects took zero scattering as the lowest category and anchored their response scales at this end.

The ends of most series of stimuli presented for judgment are not so salient as a definite zero point, but end anchoring is always a possibility when the end stimuli are easily identified as such.

Some scales may be anchored at the center when a center stimulus is easily identified as such. When a man judges the height of another man, he may take his own height as average or medium and organize his scale upward and downward from this fixed point. In one experiment on estimating heights on a nine-point scale, 152 tall men centered their scales around a mean of 69.01 inches, while 118 short men centered theirs around a mean of 68.56 inches (Hinckley and Rethlingshafer, *1951*).

A good illustration of natural anchoring comes from a study of reproductions of angular inclinations of lines, carried out by Kaufman, Reese, Volkmann, and Rogers (*1947*). The subjects were asked to adjust a line to 0, 1, 3, 6, 10 . . . degrees from the vertical. The response scale is thus a familiar one, and the responses corresponded closely to the requested stimulus values. The variability of the responses, as shown in Figure 10.17, was particularly low at and near inclinations of 0 and 90 degrees from the vertical. We may suppose that the response scale was anchored at these easily located points. Constant errors, much less predictable, are graphed in a different way in Figure 10.10. Similar results are obtained when various inclinations of lines were presented and the subjects asked to estimate the angles in degrees (Rogers, Volkmann, Reese, and Kaufman, *1947*). The same effect of anchoring at the vertical and the horizontal has been demonstrated by Chapanis using the method of reproduction (*1951a*) and the method of estimation (*1951b*). A consequence of natural anchoring is a reduction in variability of judgment at and near the anchoring stimulus. In Philip's data of Table 10.3 there is no sign of end anchoring; the 11 standard deviations are practically equal.

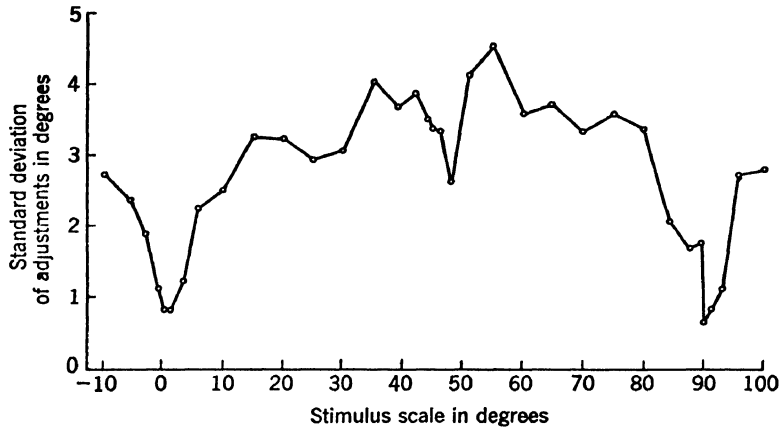


FIGURE 10.17. Effects of natural anchors at the vertical and the horizontal. Variability of adjustment is small at 0° and 90° because the response scale is tied to the stimulus scale at these points. Same data as for Figure 10.10; from Kaufman, Reese, Volkmann, and Rogers (*1947*).

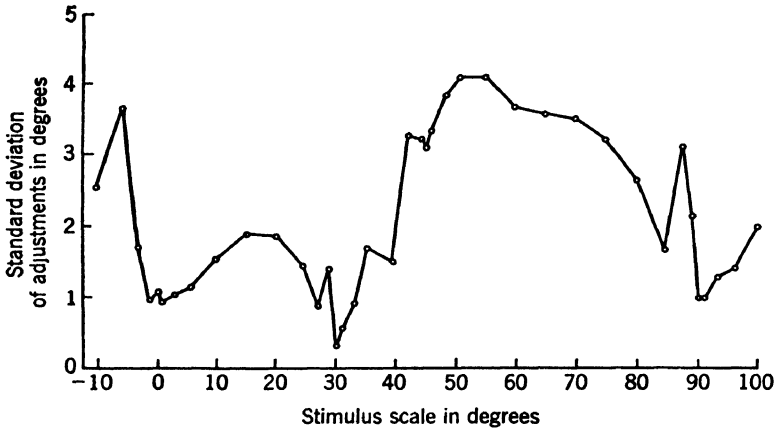


FIGURE 10.18. Effects of anchor introduced at 30°. In comparison with Figure 10.17, variability is greatly reduced near the anchor. Six subjects; 3445 judgments. From Kaufman, Reese, Volkmann, and Rogers (*1947*).

Figure 10.18 shows the effect of a *supplied* anchor. A line like the one adjusted by the subjects was presented, and the subjects were told that this line was 30 degrees. With this help the variability of adjustments at and near 30 degrees was drastically reduced. The anchoring produced by guide lines is of considerable practical importance for the design of instruments that have to be adjusted and dials that have to be read quickly and accurately.

Ratings are often made with standards or models in view, as in grading fruit and rating handwriting. These standards act as anchoring points, tieing down the scale at several places along the stimulus continuum. It is expected that such multiple anchoring will reduce the variability of judgment. Blumenfeld (*1931*) compared the judgments which result when such standards are given as types, each one representing the type or center of a category, and when they are given as limens or boundaries between categories. His results led him to favor the use of standards between categories rather than at category centers, but the evidence is not clear-cut. In the light of the prevailing custom of designing rating scales with descriptive phrases at the centers of the categories, this is still an important practical question.

Much more dramatic anchoring effects can be produced if an anchoring agent outside the stimulus series presented for judgment is supplied. Rogers (*1941*) showed his subjects lines at inclinations of 10, 16, 22, 28, 34, and 40 degrees from the vertical, to be judged on a six-point scale. The subjects soon adapted to the task and their response scales corresponded roughly with the stimulus values. Then an anchoring stimulus was supplied in the form of an angle of 40 degrees presented before each angle to be judged. The subjects were instructed that this angle was a *six*. Later other anchors were used: 45, 50, 60, 70, 80, 90, 100, and 110 degrees. The results, some of which are shown in Figure 10.19, were clear-cut. As the anchoring stimulus is extended, the response scale is extended. With a 40-degree anchor the middle category limen, i.e., the limen between Categories 3 and 4, is at 24 degrees. With a 50-degree anchor it is at 27.9 degrees. With an 80-degree anchor this limen is indeterminate because Category 4 was not used. Although the stimuli remained constant, when the anchoring stimulus was high, the response scale shifted upward and all stimuli were thus put in the low categories. The subjects asked the experimenter such questions as "Did you leave out the sixes?" Similar results are obtained in weight-lifting experiments (Rogers, *1941;* Heintz, *1950*).

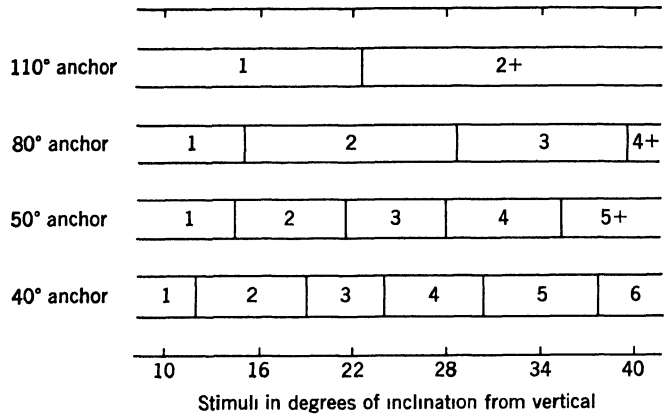


FIGURE 10.19. Effects of certain anchoring stimuli on scales of judgment. One subject made 50 judgments each of lines of 10, 16, 22, 28, 34, and 40 degrees from the vertical in terms of a six-category response scale. Just before each stimulus was presented for judgment, an anchoring stimulus was shown with instructions to call this a *six*. The response scales are systematically extended upward toward the anchoring stimulus. Data from Rogers (*1941*).

In general, as the anchoring stimulus is shifted away from the range of stimuli to be judged, the near end of the response scale shifts toward the anchoring stimulus. The far end of the scale remains anchored to the far end of the stimulus range. Since the stimuli remain the same, the far end of the scale is used more often, hence the categories at the far end become wider and the categories at the end near the remote anchoring stimulus are pushed out of use.

McGarvey (*1943*) wished to extend the study of anchoring effects to judgment of objects of more social significance than lines and weights. She used the prestige of occupations, such as seaman, bellhop, and dentist. It was necessary first to derive scale values for these occupations by a preliminary experiment. After this was done the effects of anchors at various distances (measured by these derived scale values) were studied, as by Rogers. The same procedures were followed in an experiment on judgment of the seriousness of undesirable social acts, such as "grabbing a parking space someone else is about to turn into," "putting your deformed child in the circus," and "using brass knuckles in a football game."

The results of McGarvey's experiments are quite clear-cut. "The relations found to obtain in the present experiment between the position and width of the absolute scale and the remoteness of the anchoring value are strikingly similar to those that have been demonstrated to hold for judgments of psychophysical materials." Without doubt the effects of this kind of anchoring are quite general, not limited to any particular kind of objects of judgment.

The anchoring agent need not be a stimulus that the subject actually perceives. Volkmann (*1936*) demonstrated the anchoring effect of an imagined stimulus value. His subjects judged the inclinations of lines on a six-point scale without any special anchor as a control. Then they were asked to let the horizontal define Category 1 even though no horizontal line was presented. The usual effects of anchoring appeared in the results.

Hunt and Volkmann (*1937*) had their subjects judge the pleasantness of colors on a scale of seven categories under control conditions, then each was told to select the most pleasant color he could think of and hold this color in mind as Category 7. Since this most pleasant color was more pleasant no doubt than any presented color, it acted as a remote anchor. As such it had the same effects as perceived anchoring stimuli. The scale shifted up toward it and the colors were judged lower or less pleasant than they had been in the control experiment. Hunt (*1941*) got the same results in judgments of a wide variety of stimulus objects, as varied as photographs of children and verbal statements of moral offenses.

The anchoring stimulus is more effective when it is treated as an object to be judged than when it is perceived but not judged, and a stimulus that is somewhat different from the series of objects being judged may have no anchoring effect at all unless it is also judged (Brown, *1953*).

From another point of view, namely, that of the preceding chapter, judgment of any stimulus object is influenced by the context of that object. When the effective context is a series of stimuli, a general correspondence between this series and the response scale can be observed and the experimenter can shift the response scale up or down by shifting the series of stimuli. Anchoring instructions introduce new material into the context. If the context of judgment includes remembered stimulus values or anchoring values, the experimenter cannot shift the judge's response scale so easily.

CONFIDENCE, TIME, AND DIFFICULTY

Functionally, we may say that when a person's adjustment to his situation is impeded by the complexities or indeterminateness of the situation, a judgment process is initiated. Judgment simplifies or, more precisely, categorizes the situation, permitting life to go on. Logically, we may say that the judgment settles or clarifies a problematic situation. Or, using topological concepts with Cartwright and Festinger (*1943*), we may say that when the person is in two or more equally potent overlapping situations at once, an equilibrium of forces prevents movement toward any category until the equilibrium is upset.

CONFIDENCE IN A JUDGMENT

Whichever way we say it, the number of response categories is, by definition of judgment, limited to some small number. If the judgment is made in five heterogeneous categories or on a five-point scale, there are only five responses the judge can make. But if he can respond to more than five differences in the stimulus situation, he has something left over. The five-alternative judgment does not exhaust the discriminative power of the judge, hence he will make some responses in any category faster than others in the same category and, if asked, he will report more confidence in some than in others. This variation in speed and confidence is a defining characteristic of judgment, differentiating it from uncomplicated perception and memory, though perhaps only quantitatively (see page 284).

Variations in speed and confidence also differentiate judgment from estimation (see page 284) because estimation, as in inches or ounces, uses a continuous scale with an unlimited number of response possibilities.

Reporting Confidence. Confidence or certainty of judgment is a subjective variable. All we know about anyone's confidence is what he tells us. Hence the manner of reporting confidence is quite important. The usual five-point, seven-point, and ten-point scales have been used, as well as a continuous graphic scale printed on paper which the subject checks at any point. A good argument can be made for a scale of 0 to 1 or 0 percent to 100 percent because it seems natural to the subjects. They know the meaning of zero confidence, "a pure guess," and they know the meaning of complete confidence, "100 percent certainty." Presumably both ends of the scale act as natural anchors. Frequency distributions of such confidence reports are usually J- or

U-shaped (Seward, *1928*). When these reports of confidence are added and averaged, the assumption of equivalence of units is implicit. The results obtained under good laboratory conditions by different experimenters and different subjects are regular and consistent.

Research on attitudes often makes use of a scale such as *strongly agree, agree, neutral, disagree, and strongly disagree.* It is probable that confidence in judgment is mixed up on such a scale with strength of conviction, intensity of belief, or some other oblique variable in unknown amounts. For analytical purposes a well-defined variable is much to be preferred.

THE CONFIDENCE FUNCTION. Several experimenters, from Fullerton and Cattell (*1892*) on, have asked for reports of confidence following

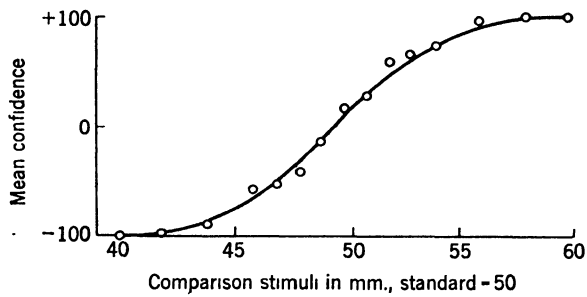


FIGURE 10.20. Confidence in the comparative judgment as a function of the difference between the stimuli. One subject judged each line ten times in relation to the standard and reported confidence on a scale of 0 to 100. When the line was judged "longer" the confidence report was given a plus sign; when it was judged "shorter" the confidence report was given a minus sign. The points are the algebraic means of the confidence reports. The curve is a normal ogive computed from these data. From Johnson (*1939*).

comparative judgments of psychophysical material. It always turns out that confidence in the two-category judgment increases with an increase in the physical difference between the variable stimulus and the standard. Johnson (*1939*) used a wide range of lines, to be judged shorter or longer, and found that the function of confidence on the stimulus variable is a normal ogive—like the usual psychometric fre-

quency function but of much more gradual slope (see Figure 10.20). This relation has been verified by Festinger (*1943a, 1943b*) for judgments of lengths of lines, but there is no reason a priori to limit the findings to this particular type of stimulus material.

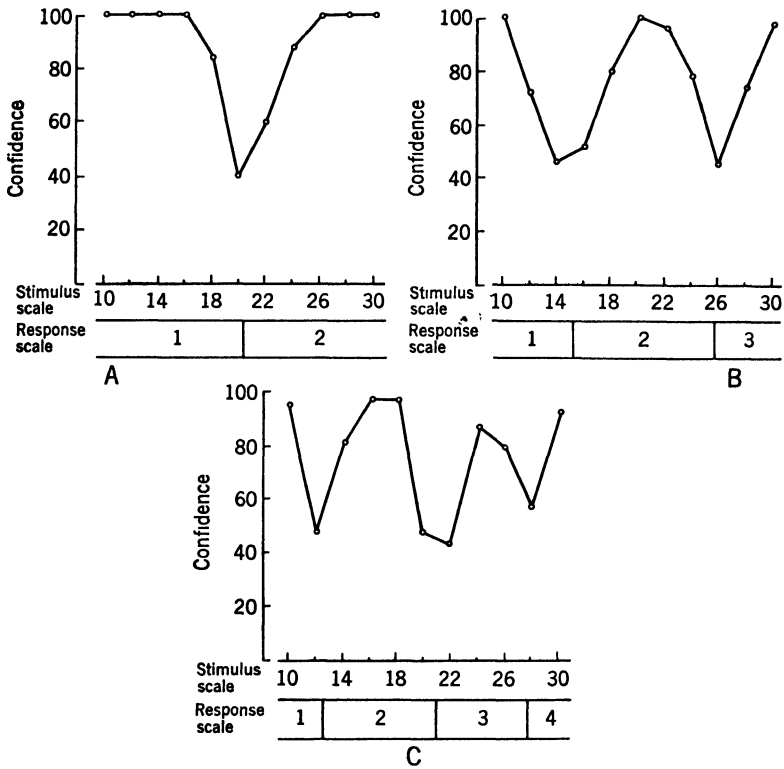


FIGURE 10.21. Confidence and Category Thresholds. The stimuli were 11 cards bearing lines of 40 symbols: 10 to 30 x's and the remainder o's. The subject was instructed to judge each stimulus card, as to preponderance of x's, in one of two categories (Fig. A), one of three categories (Fig. B), or one of four categories (Fig. C). After preliminary practice each stimulus was judged ten times in random order. Category thresholds are shown on the response scale at the bottom of each graph. In each graph mean confidence is lowest at the category thresholds.

From results of this kind the generalization has been made that *confidence increases as a function of distance from a category threshold* (Johnson *1939, 1945*). The distance involved in this generalization is,

of course, distance on the response scale, which, as noted earlier in this chapter, is functionally related to distance on the stimulus scale. But, because of the possibility of constant error, distance from the category threshold may not be exactly the same as distance from the standard stimulus.

When judgment is made in several categories and the stimulus variable is continuous, one would expect that minimum confidence would occur at each category limen and maximum confidence in the middle of each category—except at the two end categories where maximum confidence would be at the extremes. The above generalization should then be extended to read that confidence increases as a joint function of distance from the two nearest category thresholds. Recent data confirming this generalization for scales of 2, 3, and 4 categories are shown in Figure 10.21.

No doubt there are other factors in addition to the stimulus material which affect confidence in a judgment. The atmosphere of the laboratory and social relations between experimenter and subject may be mentioned as possible influences on confidence in laboratory judgments. There must be similar influences affecting extra-laboratory judgments, but no data are available on this point.

The subject's set for the judgment has been varied by instructions which emphasize speed or accuracy. The effects of such instructions on confidence are small and variable. In experiments on judgments of weights (Garrett, *1922*) and of lines (Festinger, *1943a;* Johnson, *1939*) the differences did not prove to be significant.

TIME OF JUDGMENT

Judgments may be timed easily and accurately. It is common practice to use an exposure device wired in series with an electric clock and the subject's response key, so that the clock starts when the stimulus material is exposed and stops when the subject expresses his judgment, either by speaking into a voice key or by pressing a telegraph key. As distinguished from the reaction-time experiment, the judgment-time experiment employs tasks of a greater range of difficulty and the emphasis is usually on accuracy rather than speed. The times, of course, are much longer; highly accurate timing is not necessary. It is possible, also, to use a card-sorting task to get measures of judgment time (Reed, *1951*).

THE TIME FUNCTION. In the case of the two-category judgment where various stimuli are compared with a standard stimulus, it has long been known that judgment time increases as the difference between the stimuli decreases. Figure 10.22 shows that judgment time

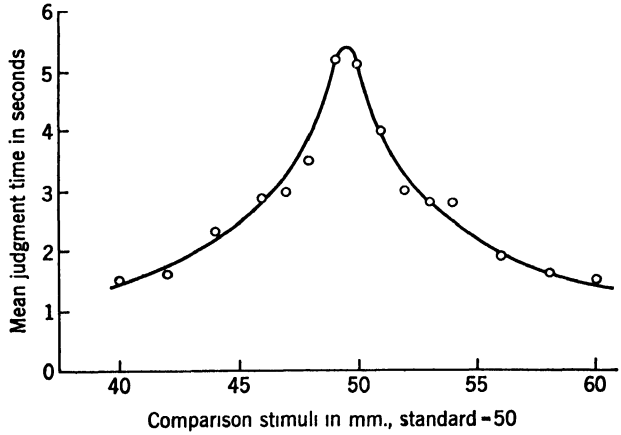


FIGURE 10.22. Judgment time in the comparative judgment as a function of the difference between the stimuli. The same 150 judgments were used from which the confidence function of Figure 10.20 was computed. From Johnson (*1939*).

reaches a peak at the point of zero stimulus difference and declines asymptotically on both sides as stimulus difference increases. Recently Reed (*1951*) has shown that the data on sorting times can be described by a hyperbolic function,

$$t = f\frac{1}{d} - k, \tag{15}$$

in which d is the stimulus difference without regard to sign and k is the theoretical judgment time for infinitely large stimulus differences. Equation 15 describes a curve that rises rapidly to a peak as stimulus difference decreases, and is thus in agreement with other published data (Henmon, *1906;* Johnson, *1939;* Cartwright, *1941a;* Festinger, *1943b*). Reed noticed that k has the same value for judgments of area

and of hue even though other constants change. Under some conditions k would be the conventional reaction time.

But judgment time by this equation would reach infinity at zero stimulus difference. Actually all published curves rise rapidly toward the point of zero stimulus difference, but round off before reaching it. People do not deliberate forever, even when two stimuli appear equal. They make a response after a while, even if only an arbitrary one. Something else, perhaps a "giving up" process, must be considered.

The effect of constant errors on the time function is also difficult to determine. By analogy with the confidence function one would expect the peak of the time function to come at the point of subjective equality or the category limen rather than at the point of objective equality or zero stimulus difference. The available data, however, do not unequivocally permit this conclusion (Johnson, *1939;* Cartwright, *1941a;* Festinger, *1943a, 1943b*). The time function may be more sensitive to constant errors than the confidence function. In any case the effect of a constant error is usually small.

When no standard stimulus is used, as in the method of single stimuli, the question of a constant error does not arise. The generalization can be made, then, that *judgment time decreases as a function of distance from a category threshold* (Johnson, *1939, 1945*). The relation is a general one that holds for scales of affective value as well as physical value (Chant and Salter, *1937;* Dashiell, *1937;* Barker, *1942, 1946;* Shipley, Norris, and Roberts, *1946*).

An extension of this generalization to judgment on a scale of several categories would state that judgment time decreases as a joint function of distance from the nearest category thresholds. Recent data confirming this generalization for scales of 2, 3, and 4 categories are shown in Figure 10.23.

Kellogg (*1931*), reporting an experiment on comparison of brightnesses viewed side by side, plotted time of judgment on the stimulus variable for all three categories of response: *left, equal, right.* He noted that the time curves tended to be inversions of the frequency curves and made the generalization from his results that "within any given category of judgment, the more the stimuli deviate from the ideally perfect condition typified by the category name, the longer the judgment-times tend to become." The generalization above is almost synonymous with Kellogg's generalization, but is stated in relation to category thresholds, which are easy to compute.

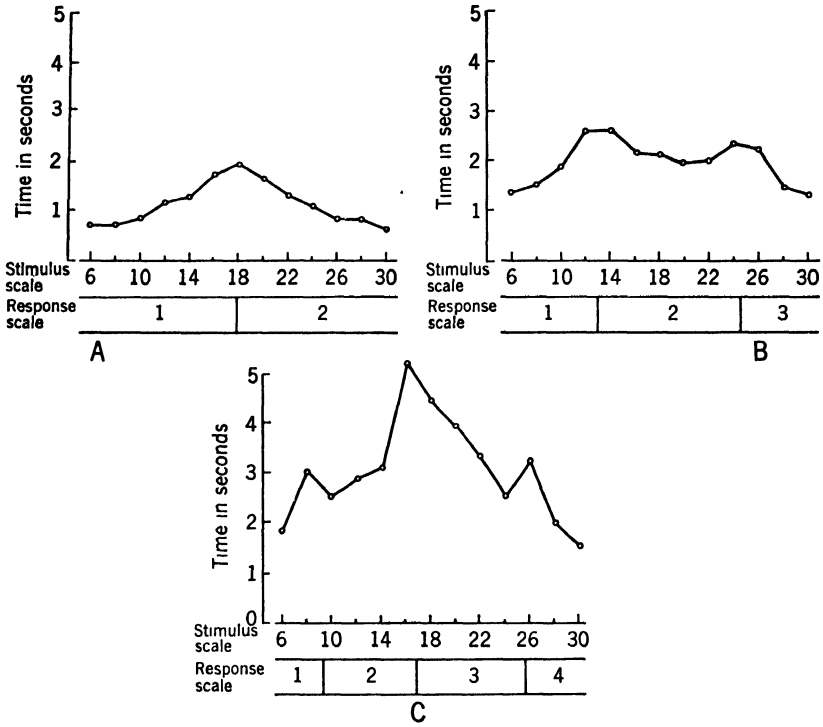


FIGURE 10.23. Judgment Time and Category Thresholds. The stimuli were 13 cards bearing 36 dots: 6 to 30 red dots and the remainder black. The subject was instructed to judge each stimulus card, as to redness, in one of two categories (Fig. A), one of three categories (Fig. B), or one of four categories (Fig. C). After preliminary practice each stimulus card was judged ten times in random order, and the data points are geometric means of these ten judgment times. Category thresholds are shown on the response scale at the bottom of each graph. In each graph judgment time approaches a peak at each category threshold.

Cartwright (*1941a*) investigated decision time under a variety of conditions. In one experiment the subjects were first shown five angles, of 60, 70, 80, 90, and 100 degrees, and told to remember them. Then they were shown many angles, of 10, 20, 30 . . . 160 degrees, one at a time and asked whether each had been presented in the learning

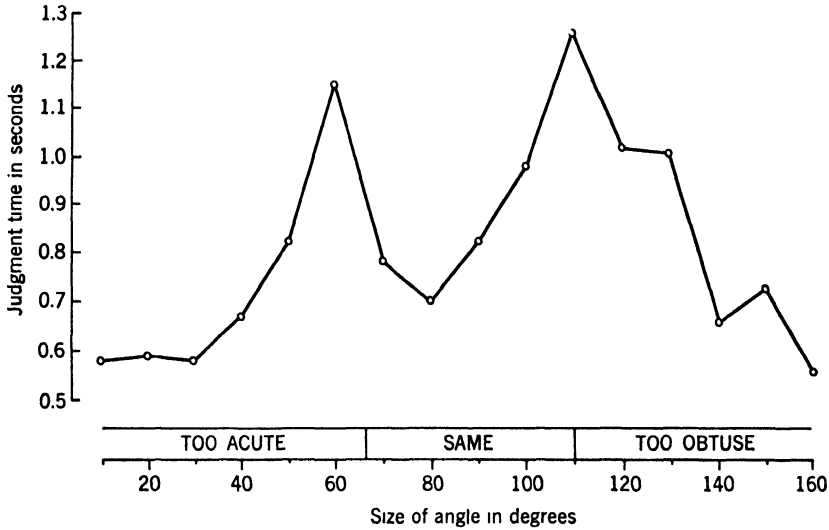


FIGURE 10.24. Judgment time for judgments of recognition. Five angles, from 60 to 100 degrees, were presented for learning. Then 16 angles, 10 to 160 degrees, were presented for judgments of recognition. Tabulation of the judgments permitted the inference that the subject had in effect divided the series of angles into three categories, the boundaries of which were computed as 66 and 105 degrees. Judgment time was maximal at these two boundaries. From Cartwright (*1941a*).

series. Thus the subjects made their decisions in two categories, but the results gave evidence that the scale of angles from 10 to 160 degrees was divided into three "ranges of equivalence," which in this experiment might be called *too acute, same, too obtuse.* Therefore there were two category thresholds or borders between the "ranges of equivalence," which could be calculated from the obtained frequencies of each judgment. As one would expect, decision time was maximal at the two thresholds (see Figure 10.24). The two peaks are very definite and each is located at a category threshold.

From these relations it is a logical deduction that if the middle category were much narrower, there would be no room for the time curve to dip between the peaks, hence only one peak would be obtained. Cartwright proved this deduction by presenting only one angle, of 90 degrees, in the learning series, then testing as before. The results

show a very narrow middle category and a time curve of only one peak.

As Cartwright states the general principle, "the time required for deciding to which category a stimulus belongs increases as the stimulus approaches the border of a range of equivalent stimuli." This principle, in whatever language one chooses to put it, can now be considered well established. Some of the conditions, of instruction, of context, and the like, which determine the width of categories or of ranges of equivalence have been discussed earlier.

Like confidence in judgment, time of judgment must be affected by other factors in the situation. In the laboratory, for example, urgent, hasty actions by the experimenter would probably speed up the subject's judgments, but there are no data on this conjecture. Instructions emphasizing speed or accuracy do produce large differences in time of judgment (Johnson, *1939;* Festinger, *1943b*). Practice in judging reduces judgment time, even though the stimulus material remains approximately constant (Gallenbeck and Smith, *1950*). Judgments are slower when the judge sees an audience watching him than when he sees the experimenter only, and even slower when he knows that an audience is watching him behind a one-way mirror (Wapner and Alper, *1952*).

DIFFICULTY OF JUDGMENT

There are four concepts of difficulty applicable to the difficulty of judgments: felt difficulty, effort expended, percentage of errors by an individual, and percentage of errors by a group.

Felt difficulty is of course a subjective variable. One can get at it only by asking a judge to rate the difficulty of a judgment just made. In all probability, ratings of the difficulty of judgments and of confidence in these judgments would show a high negative correlation. These two terms may be antonymous designations for the same subjective variable. Hence the information on confidence summarized above applies also to felt difficulty.

When motivation is constant, people will work harder on difficult judgments than on easy ones. The relation is not linear, however, because when the difficulty is so overwhelming that extra effort is unrewarding, people simply give up. Up to this point physiological signs of effort expended may be taken as measures of difficulty.

When the judgments can be objectively marked right or wrong,

those stimulus objects are most difficult for any subject to judge which he judges wrong most often. This truism can be put to work measuring the difficulty of judgments that can be repeated often enough to yield a stable percentage of wrong judgments.

Judgments of interesting complex objects can be remembered, but judgments of lines, weights, brightnesses, and the like can be repeated, as in psychophysical experimentation, without fear of memorizing. Comparative judgment in two categories of such stimulus material yields a measure of difficulty in this sense which is easily related to the physical stimulus variable and thus yields a difficulty function. In fact the usual psychometric frequency function can be read so as to yield

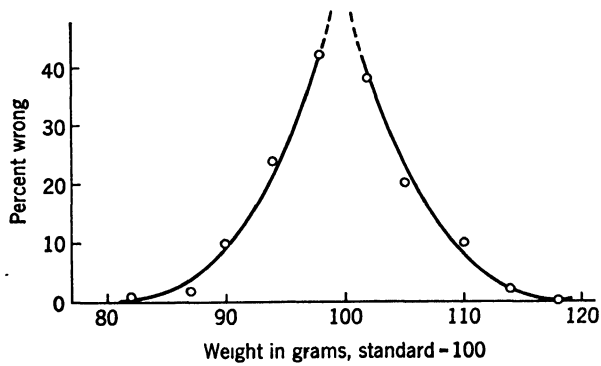


FIGURE 10.25. Difficulty of judgment in terms of percent wrong as a function of stimulus difference in a weight-lifting experiment.

difficulty values. The frequency function, graphing the percentage of judgments in each category, e.g., *longer* and *shorter,* covers the transition zone from 100 percent in one category to 100 percent in the other. At the transition point or category limen, the stimulus objects would be put in each category equally often—ignoring constant errors for the moment—hence half the judgments are wrong. At this point on the stimulus variable, judgment is maximally difficult. The difficulty of the judgments increases as the stimuli near the category limen or point of subjective equality. Difficulty decreases, in this sense of percentage of wrong judgments, at a decreasing rate, approaching zero percent wrong asymptotically in both directions (see Figure 10.25).

The most common measure of difficulty in the construction of tests

of abilities and achievements is a group measure, namely, the percentage of a number of subjects who judge an item erroneously. The uses and limitations of this technique are well known. If group difficulty values were determined and plotted on the stimulus variable, the relations would be similar to those discussed above for the individual subject—with the proviso that they might not hold in detail for any particular subject.

One would expect that these different measures of difficulty of judgment are related, and fortunately there is some evidence that this is true. Felt difficulty is related to objective stimulus difference and to the logarithm of scale values determined by group percentage wrong (Guilford and Cotzin, *1941*). Percentage wrong is related to an increase in muscular activity (Davis, *1938*) and a drop in palmar skin resistance (Freeman and Giese, *1940*). If these studies of difficulty were extended to the case of judgment in several categories, the difficulty curves would no doubt resemble the time curves of Figure 10.23.

RELATIONS BETWEEN DEPENDENT VARIABLES

We have just seen that the minor dependent variables of judgment, confidence, time, and difficulty are related to the major dependent variable, frequency of use of each response category. These minor variables are also related to each other in a systematic way.

When anyone judges many similar objects, the time taken for each judgment will be inversely related to the confidence reported for each. Seward (*1928*) studied this relation by the correlation technique, using recognition judgments of fancy papers, some of which had been shown to the subjects previously and some of which were new. Her correlations for many subjects ranged from .37 to .81, with a median of .65. The correlation technique assumes a linear relationship, to be sure. Two attempts have been made to determine the form of the relationship empirically. Volkmann (*1934*), on the basis of judgments of inclinations of lines, published a tentative equation in the form of a hyperbola. Johnson's (*1939*) equation, which fits the data on judgments of lengths of lines, recognition of figures, and meanings of words fairly well, is in logarithmic form: "As doubt or uncertainty increases arithmetically, judgment time increases geometrically." The departures from these simple equations are chiefly in the region of the threshold where the large degree of doubt may introduce a complication in the manner of reaching the decision, e.g., "giving up."

Since confidence decreases and felt difficulty increases as a category threshold is approached, one would expect that, when many judgments of one kind of stimulus material are made, these two variables would be inversely related. This expectation has been verified by Farmer (*1928*) and Hertzman (*1937*). In general, a change in the stimulus material or conditions of judgment which increases accuracy of judgment also increases confidence in the judgment (Garrett, *1922;* Greene, *1929;* Johnson, *1939*). Learning experiments have been arranged so that confidence and percentage of correct judgments rise in a more-or-less parallel course (Johnson and Rhoades, *1941*). Under some conditions the relationship may be reversed: those judgments given with the most confidence are the most accurate (Cady, *1923;* Öbrink, *1948*). But when the stimulus material is not homogeneous, confidence of judgment cannot be taken as a guide to the correctness of the judgment (Kelly and Fiske, *1951*).

Though there seems to be no evidence at hand, we may be sure that time of judgment and difficulty of judgment will be related, however difficulty is measured.

In general, then, if someone is asked to judge a homogeneous series of stimulus objects, as the stimulus objects approach a category limen of the response scale, the difficulty of the judgments increases, time of judgment increases, and confidence in the judgments decreases. Regardless of the stimulus dimension, time and difficulty of judgment will be positively related and both will be negatively related to confidence of judgment.

Some Consequences of These Relationships. For some purposes these relationships can be ignored or the constant errors they introduce assumed to be randomized. But when conclusions are drawn from comparisons of different kinds of judgments, fallacious interpretations are possible. It may happen, for example, that one category of judgment is given with more confidence, conviction, or intensity, or less time or greater effort, than another, but the difference cannot be adequately interpreted unless distance from the category threshold is considered. If wrong judgments are given with more confidence than right judgments, it may simply be because they are more distinctly wrong; i.e., the items judged *wrong* are farther from the limen between *right* and *wrong* than the items judged *right.* A study of the relationship between atypical opinion and confidence will illustrate the value of this kind of analysis.

The question was first studied by Allport and Hartman (*1925*). They obtained opinions on controversial topics and reports of confidence in these opinions. The results indicated that those who held extreme opinions (extremely radical or extremely conservative) were more confident. The extremist's excess of confidence was attributed to compensation for a feeling of insecurity in his position.

The question was reëxamined (Johnson, *1940*) with the aid of the Thurstone scales of attitude toward war and censorship. In a sample of 68 college students no extremely prowar or procensorship students

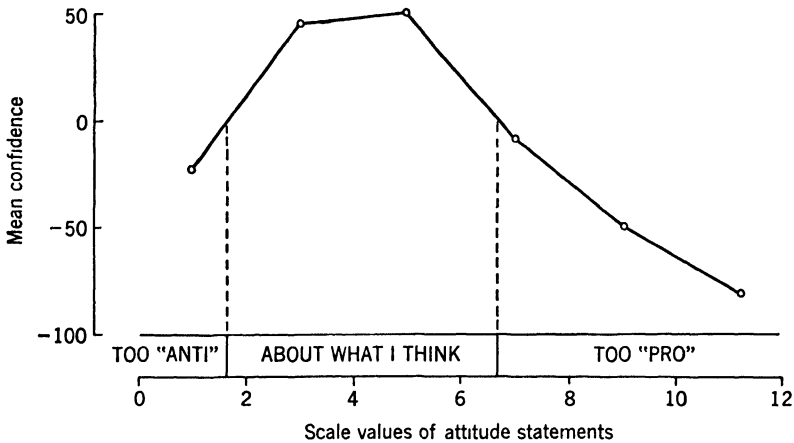


FIGURE 10.26. Relation between confidence and extremeness of attitude. The subject judged 84 statements about war taken from Thurstone attitude scales and reported his confidence in each judgment. Confidence reports were averaged algebraically, as for Figure 10.20. The highest confidence comes from rejection of statements far from a category boundary. Data from Johnson (*1940*).

were found, but the antiwar and anticensorship extremists were clearly more confident in expressing their opinions than those in the middle of each scale. The relation between atypicality of opinion and confidence is thus verified—at the "anti" end of each scale at least—but the reason for the relation can be clarified by studying one subject's judgments in more detail. Figure 10.26 shows confidence in accepting and in rejecting 84 statements at various points on the scale of attitude toward war. This subject's attitude, as measured by the median value of the statements accepted, was 4.3, and the graph shows that he divided

the scale, in effect, into three categories. Statements having scale values between 0 and 1.8 were "too anti." Those between 1.8 and 6.4 were "acceptable," or "about what I think." Those above 6.4 were "too pro." Separating these three categories are two category limens at which confidence is minimal. Confidence is maximal in rejecting statements at the "pro" end because it is here that statements are farthest from a category limen. Highest confidence therefore comes from negative judgments of statements far from the subject's position on the attitude continuum.

It is conceivable that the middle category, containing statements that are judged acceptable, could be a wide one. Then high confidence would come from affirmative judgments of statements near the subject's position. But this is not usually the case in attitude scales, nor in common public expressions of opinion. High confidence comes more often from negative judgments.

Now it is obvious that if the subject is at either extreme of the scale, he will be able to give many negative judgments, of statements at the other extreme, with high confidence. The perplexed middle-of-the-roader will find few statements so far from his position that he can reject them confidently. Thus, in place of an assumed relation between atypicality of opinion and confidence, we arrive at a demonstrable relation consequent to the proved generalization about confidence and distance from category limens. Similar analyses of effort and time of judgment would yield similar conclusions.

The relation between judgment time and distance from the category threshold has important consequences also when the two-category judgment is compared with the three-category judgment. Kellogg (*1931*), working with brightness judgments, found that judgments in three categories, one of which was the *equal* category, required on the average about 10 percent more time than judgments of the same stimuli in two categories. This finding might be taken as evidence that the equal judgment is intrinsically more difficult, less natural. A more convincing explanation is that in the three-category judgment there are two category thresholds nearness to which slows judgment, while in the two-category judgment this retardation occurs at only one threshold. If the stimuli are continuously and uniformly located along the stimulus variable, this difference will always be found. Under these conditions the generalization may be stated: *average judgment time and difficulty decrease, and confidence increases,*

with an increase in the number of categories on the response scale. In topological language the generalization was stated some years ago by Cartwright (*1941b*): decision time increases with an increase in differentiation of the phenomenal field. The evidence in support of this generalization comes only from a comparison of two-category and three-category judgments. Nothing is known of the upper limit.

INDIVIDUAL DIFFERENCES

Regardless of the accuracy of judgment, individual differences in the use of a scale of judgment have often been noticed. Several experimenters have suggested, for example, that some judges tend to use the high end of a scale, while others use the low end. Some prefer the extreme categories, while others prefer the pallid mean. In the apocrypha of psychology it is asserted that a collection of ratings tells as much about the rater as it does about the objects being rated. The first question, of course, is whether people are consistent in these respects.

This question of consistency or reliability can be attacked by statistical analysis of ratings.[3] Twenty-six college girls rated 32 girls living in the same sorority house on a seven-point scale in respect to sociability. When the ratings made by each judge are split in half, we may compare the mean rating of 16 girls by each judge with the mean rating of the other 16 girls by the same judge. The correlation between these two distributions of means for the 26 judges was .93. Unquestionably those who rated one half of the girls high rated the other half high.

A corresponding correlation for the dispersions of the ratings can be obtained in the same way by first computing the standard deviation of 16 of the judgments made by each judge and the standard deviation of the other 16 judgments by the same judge. In this case the standard deviations ranged from .96 to 2.93. The correlation between the two standard deviations for the 26 judges was .64. Again an appreciable internal consistency is demonstrated. If we apply the Spearman-Brown formula to these coefficients to estimate the reliability of scores based on all 32 judgments, we get .96 for the reliability of the means and .78 for the reliability of the dispersions.

Some evidence on the repeat reliability of such aspects of judging behavior comes from a recent study by Berg and Collier (*1953*). Sixty abstract designs were judged on a four-category scale: *like much,*

[3] These ratings were collected for another purpose by Robert J. Barndt.

like slightly, dislike slightly, dislike much. It is hardly necessary to calculate standard deviations for such a short scale. A score for extremism, or the tendency to use the extreme categories, was calculated for each judge by simply counting the number of extreme judgments. In several different groups retest reliability coefficients over periods of 4 to 15 days were in the neighborhood of .80. When similar judgments of 60 words were collected from four groups of subjects at the same time as the judgments of 60 designs, frequency of extreme judgments on the two tasks correlated between .34 and .77 (Berg, *1953*). Frequency of extreme judgments on personality-test items judged in five categories and interest-test items judged in five categories correlated .40 (Rundquist, *1950*).

The next question concerns the relation between individual differences in the use of a scale and other individual differences. In one questionnaire study that asked for opinions on psychotherapy, psychoanalyzed therapists expressed their opinions in extreme categories much more frequently than nonanalyzed therapists (Seeman, *1950*). In the Berg and Collier study, college women used extreme categories significantly more than college men, Negro students more than white students, and high-anxiety students more than low-anxiety students.

People do not always continue to use the same scale of judgment. When stimulus objects of higher or lower magnitude are presented for judgment and the response scale shifts up or down, dominant subjects seem to shift more readily than submissive subjects (Tresselt and Becker, *1950*).

It is possible to use the anchoring experiment to get scores indicating susceptibility to the anchoring effect. Frawley (*1948*) used 15 series of judgments in this way, made up from the Thurstone attitude scales. Intercorrelations of the 15 anchoring scores for 115 students were practically all positive. Factor analysis disclosed a general factor and group factors presumably due to the content of some of the attitude scales. Apparently there is some generality in the anchoring effect. The general factor was not related to conventional measures of intelligence or reasoning.

These experiments on individual differences in the use of a scale of judgment do not go far, but as far as they go they indicate that reliable scores can be obtained and that these scores are worth further investigation as objective measures of personality or style of judgment.

Individual differences in confidence in judgment are easily and

reliably measured. When different kinds of material are presented for judgment, the amount of generality under favorable conditions is indicated by intercorrelations in the neighborhood of .50 (Johnson, *1939, 1940;* Öbrink, *1948*). Bifactor analysis of confidence scores for judging the items on eight vocabulary tests allotted 48 percent of the variance to the general factor (Johnson, *1941*). Ego involvement seems to be a factor in these correlations. When Klein and Schoenfeld (*1941*) emphasized the personal significance of the judgments, mean intercorrelations of confidence scores was .45. When there was no such emphasis, the corresponding mean was .22.

When students deliberate over the items on an ordinary achievement test, individual differences in confidence are not closely related to intelligence or to performance (Trow, *1923;* Greene, *1929;* Johnson, *1941;* Öbrink, *1948*). Confidence enters the picture in another way, however. If students taking a multiple-choice test are told to answer only the questions they are sure of, i.e., not to guess, the more confident ones will attempt more items and thus will get more right answers (Gritten and Johnson, *1941*). The conventional correction—which might be called a correction for variations in confidence—is fairly adequate for the removal of this advantage.

Since judgment is slow, impeding the "congenial pace" of the individual, one would expect that some people would be more eager, as a general rule, to bring judgment to a close than others. When the judgments are carefully timed, intercorrelations are high (Johnson, *1939*).

As mentioned above, any one individual is likely to judge those items with confidence that he judges quickly. This is an ipsative or intra-individual correlation. Inter-individual correlations are much lower. That is, when different people are compared, an individual who judges many items with higher-than-average confidence will not necessarily judge them with higher-than-average speed (Johnson, *1939*).

REFERENCES

Allport, F. H., and Hartman, D. A., 1925. The measurement and motivation of atypical opinion in a certain group, *Am. Polit. Sci. Rev., 19:* 753–763.

Barber, J. L., and Garner, W. R., 1951. The effect of scale numbering on scale-reading accuracy and speed, *J. Exper. Psychol., 41:*298–309.

Barker, R. G., 1942. An experimental study of the resolution of conflict by

children, in *Studies in personality,* ed. Q. McNemar and M. A. Merrill, New York, McGraw-Hill Book Co.

Barker, R. G., 1946. An experimental study of the relationship between certainty of choice and the relative valence of the alternatives, *J. Personality, 15:*41–52.

Berg, I. A., 1953. The reliability of extreme position response sets in two tests, *J. Psychol., 36:*1–9.

Berg, I. A., and Collier, J. S., 1953. Personality and group differences in extreme response sets, *Educ. Psychol. Meas., 13:*164–169.

Blumenfeld, W., 1931. Urteil und Beurteilung, *Arch. f. d. ges. Psychol.,* Ergänzungsband 3.

Bressler, J., 1933. *Judgment in absolute units as a psychophysical method,* Arch. Psychol., No. 152.

Brotherton, D. A., Read, J. M., and Pratt, K. C., 1948. Indeterminate number concepts: II. Application by children to determine number groups, *J. Genet. Psychol., 73:*209–236.

Brown, D. R., 1953. Stimulus-similarity and the anchoring of subjective scales, *Am. J. Psychol., 66:*199–214.

Cady, V. M., 1923. *The estimation of juvenile incorrigibility,* J. Delinq. Monogr., No. 2.

Cartwright, D., 1941a. Relation of decision-time to the categories of response *Am. J. Psychol., 54:*174–196.

Cartwright, D., 1941b. Decision-time in relation to the differentiation of the phenomenal field, *Psychol. Rev., 48:*425–442.

Cartwright, D., and Festinger, L., 1943. A quantitive theory of decision, *Psychol. Rev., 50:*595–621.

Chant, S. N. F., and Salter, M. D., 1937. The measuring of attitudes toward war and the galvanic skin reflex, *J. Educ. Psychol., 28:*281–289.

Chapanis, A., 1951a. Studies of manual rotary positioning movements: I. The precision of setting an indicator knob to various angular positions, *J. Psychol., 31:*51–64.

Chapanis, A., 1951b. Studies of manual rotary positioning movements: II. The accuracy of estimating the position of an indicator knob, *J. Psychol., 31:*65–71.

Dashiell, J. F., 1937. Affective value-distances as a determinant of esthetic judgment-times, *Am. J. Psychol., 50:*57–67.

Davis, R. C., 1938. The relation of muscle action potentials to difficulty and frustration, *J. Exper. Psychol., 23:*141–158.

Farmer, E., 1928. Concerning subjective judgment of difficulty, *Brit. J. Psychol., 18:*438–442.

Fernberger, S. W., 1930. The use of equality judgments in psychophysical procedures, *Psychol Rev., 37:*107–112.

Festinger, L., 1943a. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference, *J. Exper. Psychol., 32*:291–306.

Festinger, L., 1943b. Studies in decision: II. An empirical test of a quantitative theory of decision, *J. Exper. Psychol., 32*:411–423.

Frawley, P. J., 1948. A study of judgment: a factorial analysis of the anchoring effects, Washington, D. C., Catholic University of America Press.

Freeman, G. L., and Giese, W. J., 1940. The relationship between task difficulty and palmar skin resistance, *J. Gen. Psychol., 23*:217–220.

Fullerton, G. S., and Cattell, J. Mck., 1892. On the perception of small differences, *Publ. Univ. Penn. Philos. Series,* No. 2.

Gallenbeck, C., and Smith, K. U., 1950. Systematic formulation and experimental analysis of the phenomena of thinking and belief, *J. Exper. Psychol., 40*:74–80.

Garrett, H. E., 1922. *A study of the relation of accuracy to speed,* Arch. Psychol., No. 56.

Greene, E. B., 1929. Achievement and confidence on true-false tests of college students, *J. Abnorm. & Social Psychol., 23*:467–478.

Gritten, F., and Johnson, D. M., 1941. Individual differences in judging multiple-choice questions, *J. Educ. Psychol., 32*:423–430.

Guilford, J. P., 1936. *Psychometric methods,* New York, McGraw-Hill Book Co.

Guilford, J. P., and Cotzin, M., 1941. Judgment of difficulty of simple tasks, *Am. J. Psychol., 54*:38–52.

Heintz, R. K., 1950. The effect of remote anchoring points upon the judgment of lifted weights, *J. Exper. Psychol., 40*:584–591.

Helson, H., 1947. Adaptation-level as frame of reference for prediction of psychophysical data, *Am. J. Psychol., 60*:1–29.

Helson, H., 1948. Adaptation-level as a basis for a quantitative theory of frames of reference, *Psychol. Rev., 55*:297–313.

Henmon, V. A. C., 1906. *The time of perception as a measure of differences in sensations,* Arch. Phil., Psychol., and Sci. Methods, No. 8.

Hertzman, M., 1937. Confidence ratings as an index of difficulty, *J. Exper. Psychol., 21*:113–119.

Hinckley, E. D., and Rethlingshafer, D., 1951. Value judgments of heights of men by college students, *J. Psychol., 31*:257–262.

Hollingworth, H. L., 1909. *The inaccuracy of movement,* Arch. Psychol., No. 13.

Hollingworth, H. L., 1910. The central tendency of judgment, *J. Philos., Psychol. & Sci. Method., 7*:461–469.

Hunt, W. A., 1941. Anchoring effects in judgment, *Am. J. Psychol., 54:* 395–403.

Hunt, W. A., and Volkmann, J., 1937. The anchoring of an affective scale, *Am. J. Psychol., 49:*88–92.

Ipsen, G., 1926. Über Gestaltaufassung. Erörterung des Sanderschen Parallelogramms, *Neue psychol. Stud., 1:*167–278.

Johnson, D. M., 1939. *Confidence and speed in the two-category judgment,* Arch. Psychol., No. 241.

Johnson, D. M., 1940. Confidence and the expression of opinion, *J. Social Psychol., 12:*213–220.

Johnson, D. M., 1941. Confidence and achievement in eight branches of knowledge, *J. Educ. Psychol., 32:*23–36.

Johnson, D. M., 1944. Generalization of a scale of values by the averaging of practice effects, *J. Exper. Psychol., 34:*425–436.

Johnson, D. M., 1945. A systematic treatment of judgment, *Psychol. Bull., 42:*193–224.

Johnson, D. M., 1946. How a person establishes a scale for evaluating his performance, *J. Exper. Psychol., 36:*25–34.

Johnson, D. M., 1949a. Generalization of a reference scale for judging pitch, *J. Exper. Psychol., 39:*316–321.

Johnson, D. M., 1949b. Learning function for a change in the scale of judgment, *J. Exper. Psychol., 39:*851–860.

Johnson, D. M., and Rhoades, C., 1941. Measurement of a subjective aspect of learning, *J. Exper. Psychol., 28:*90–92.

Kaufman, E. L., Reese, T. W., Volkmann, J., and Rogers, S., 1947. *Accuracy, variability and speed of adjusting an indicator to a required bearing,* So. Hadley, Mass., Mt. Holyoke College.

Kellogg, W. N., 1931. Time of judgment in psychometric measures, *Am. J. Psychol., 43:*65–86.

Kelly, E. L., and Fiske, D. W., 1951. *The prediction of performance in clinical psychology,* Ann Arbor, University of Michigan Press.

Klein, G. S., and Schoenfeld, N., 1941. The influence of ego-involvement on confidence, *J. Abnorm. & Social Psychol., 36:*249–258.

McGarvey, H. R., 1943. *Anchoring effects in the absolute judgment of verbal materials,* Arch. Psychol., No. 281.

Michels, W. C., and Helson, H., 1949. A reformulation of the Fechner law in terms of adaptation-level applied to rating-scale data, *Am. J. Psychol., 62:*355–368.

Mt. Holyoke Psychophysical Research Unit, 1953. *Psychophysical research, summary report 1946–1952,* So. Hadley, Mass., Mt. Holyoke College.

Nash, M. C., 1950. An experimental test of the Michels-Helson theory of judgment, *Am. J. Psychol., 63:*214–220.

Öbrink, J., 1948. An experimental investigation of confidence, Ph. D. thesis, Uppsala, Sweden, University of Uppsala.

Philip, B. R., 1947. Generalization and central tendency in the discrimination of a series of stimuli, *Canad. J. Psychol., 1*:196–204.

Postman, L., 1946. The time-error in auditory perception, *Am. J. Psychol., 59*:193–219.

Reed, J. B., 1951. *The speed and accuracy of discriminating differences in hue, brilliance, area, and shape,* So. Hadley, Mass., Mt. Holyoke College.

Rogers, S., 1941. *The anchoring of absolute judgments,* Arch. Psychol., No. 261.

Rogers, S., Volkmann, J., Reese, T. W., and Kaufman, E. L., 1947. Accuracy and variability of direct estimates of bearing from large display screens, So. Hadley, Mass., Mt. Holyoke College.

Rundquist, E. A., 1950. Response sets: a note on consistency in taking extreme positions, *Educ. Psychol. Meas., 10*:97–99.

Seeman, W., 1950. Clinical opinion on the role of therapist adjustment in psychotherapy, *J. Cons. Psychol., 14*:49–52.

Seward, G., 1928. *Recognition time as a measure of confidence,* Arch. Psychol., No. 99.

Shipley, W. C., Norris, E. D., and Roberts, M. L., 1946. The effect of changed polarity of set on decision time of affective judgments, *J. Exper. Psychol., 36*:237–243.

Taubman, R. E., 1950. Studies in judged number: I. The judgment of auditory number, *J. Gen. Psychol., 43*:167–194.

Tresselt, M. E., 1947. The influence of amount of practice upon the formation of a scale of judgment, *J. Exper. Psychol., 37*:251–260.

Tresselt, M. E., 1948. The effect of the experiences of contrasted groups upon the formation of a new scale of judgment, *J. Social Psychol., 27*: 209–216.

Tresselt, M. E., and Becker, M., 1950. Scales of judgment and personality correlates, *J. Gen. Psychol., 43*:221–230.

Trow, W. C., 1923. *The psychology of confidence,* Arch. Psychol., No. 67.

Turchoie, R. M., 1948. The relation of adjacent inhibitory stimuli to the central tendency effect, *J. Gen. Psychol., 39*:3–14.

Volkmann, J., 1934. The relation of time of judgment to certainty of judgment, *Psychol. Bull., 31*:672–673.

Volkmann, J., 1936. Anchoring of absolute scales, *Psychol. Bull., 33*:742–743.

Volkmann, J., 1951. Scales of judgment and their implications for social psychology, in *Social psychology at the crossroads,* ed. J. H. Rohrer and M. Sherif, New York, Harper & Brothers.

Wapner, S., and Alper, T. G., 1952. The effect of an audience on behavior in a choice situation, *J. Abnorm. & Social Psychol., 47*:222–229.

Wever, E. G., and Zener, K. E., 1928. The method of absolute judgment in psychophysics, *Psychol. Rev., 35*:466–493.

Woodrow, H., 1933. Weight discrimination with a varying standard, *Am. J. Psychol., 45*:391–416.

Woodworth, R. S., 1938. *Experimental psychology,* New York, Henry Holt and Co.

Chapter 11

Individual Differences in Thought and Judgment

Style is the physiognomy of the mind, and a safer index to character than the face.

—Schopenhauer

THE individual differences most frequently considered in the psychology of thinking are scores representing abilities, i.e., measures of performance when working conditions and motivation are at an optimum. Other individual differences, such as flexibility, spontaneity, persistence, sensitivity, forethoughtfulness, comprehensiveness, pedantry, distractibility, and susceptibility to emotional distortion, have been considered less frequently but may turn out to be equally important. In general there are two large topics for this chapter: the relationships between scores representing various individual differences, and the relationships between these scores and real-life behavior. These topics have been approached in three ways.

1. As various intellectual activities have been identified and studied in the laboratory, individual differences in these activities have occasionally been studied also, e.g., individual differences in word association, judgment time, and stereotypy. Some of the results of these investigations, even though they are usually only by-products of an interest in general principles, are valuable for our purposes and have been discussed in preceding chapters, where it may be seen that most of the scores used have been rather crude: simple counts of various types of response, and simple measures of elapsed time. If the general principle can be stated in the form of a mathematical equation, in-

dividual differences appear as parameters of these equations. Such parameters can then be treated as scores, compared, and analyzed for their own sake. This elegant mathematical procedure has not often been applied to the study of individual differences—though an illustration may be found on page 206—and it may turn out that cruder scores are just as useful.

Whatever type of score has been used, when the study of individual differences is a by-product of an interest in something else, the conventional methodology of test construction and refinement has frequently been ignored. As noted in preceding discussions, questions of internal consistency and repeat reliability have often not been answered, often not even asked.

2. Tests of abilities and other individual differences have frequently been devised on the basis of educational or industrial experience with little regard to the concepts and principles flowing out of experimental psychology. In fact discussions of intellectual abilities in psychology textbooks usually have little to do with discussions of general principles of thinking. The construction of tests of general intelligence, arithmetic, vocabulary, reasoning, and mechanical aptitude illustrates this approach.

Tests developed according to this approach have in general profited by a sophisticated methodology of test construction and evaluation. Relationships between abilities have been analyzed in some detail, often, in recent years, by factor analysis. Often individual differences on these tests have been compared with an external criterion of success in school, in industry, and in military service. Information about relationships between test scores and criteria of success in specified jobs furthers description of the abilities represented by the scores. Much of the information assembled in this chapter is the result of this approach.

3. The third approach starts at the criterion. We can begin with a sample of people known to be good thinkers in some respect, e.g., recognized artists, and compare them with the ordinary run of mortals in as many different ways as possible. The study of genius illustrates this approach. The principal difficulty, as we shall see later in this chapter, is that, when we are sure of our criterion, as in the case of historical geniuses, the geniuses are not available for testing. Comparisons are frequently limited to data on record, such as ancestry, education, and notable childhood accomplishments.

GENERAL PROBLEM-SOLVING ABILITY

The attempt to test general ability to solve problems suffers, like so many other testing attempts, from the lack of a good criterion. If a group of people who are known to be all-round good problem solvers could be identified and compared with a group known to be generally bad problem solvers, the construction and evaluation of tests of general problem-solving ability could proceed directly. Actually no such criterion is available. People can be grouped according to their success or failure on specific problems or even on specific jobs that presumably require thought. Those who have demonstrated good judgment or creative accomplishment can be singled out for special study. But there are no criteria of all-round ability to think, no records that differentiate people who have solved a wide variety of problems from people who have tried such problems and failed. There is no clear evidence from records of life accomplishment for or against a general factor of ability to succeed, because the problems that life offers its competitors are not all equally difficult.

It is possible that laboratory problems could be assembled as a criterion of problem-solving ability. A wide variety of problems would be required, and of course good logic requires that this work sample be a representative sample of life's problems. Unfortunately, the population of problems from which such a sample is to be drawn is unknown.

Lacking any broad sample of problems or any defensible criterion groups, a number of investigators have done the next best thing. Ignoring the criterion problem, they have simply given their laboratory subjects several problems and have correlated achievement on one with achievement on the others. Low positive correlations are usually reported. Probably larger correlations would result if the number of problems of each kind were large enough to constitute a reliable test of each kind. No really broad sampling of problems has been reported.

An investigation by Billings (*1934*) is probably the most thorough attempt to study individual differences in solving problems of different kinds. (Also see page 139.) And these are limited to academic problems that require reading. He gave 60-minute tests of problem solving in eight areas: geometry, arithmetic, physics, mechanics, economics, sociology, geography, and history. Since he wanted to test problem solving independent of mastery of the material of the problem, he gave his subjects instruction in the materials of the problems in advance of

the problem solving itself. Average intercorrelations of the scores obtained by this technique was .67. Correlations between the problem-solving tests and knowledge of the material of the problem in the same field were definitely lower than this. As far as these data go, they constitute a strong argument for a general factor of problem solving within the limits of this area of printed tests of scholastic problems. It is possible that there is a general reading factor or a general reasoning factor or general intelligence factor in these correlations. If a wider variety of problems, such as mechanical puzzles and problems of interpersonal relations, were included in the battery, intercorrelations would probably be lower, but there is no direct evidence on this question. It is also possible that in a broader sample of subjects the correlations would be higher.

GENERAL INTELLIGENCE AS PROBLEM-SOLVING ABILITY

A good argument could be put up for defining general intelligence as the ability to solve the general run of life's problems. Inspection of ordinary intelligence tests will show that they are composed in large part of minor problems to be solved. In part also they test abilities that contribute in a general way to problem solving, e.g., attention span.

If we take success in life as a rough index of ability to solve life's problems, the evidence of the importance of general intelligence is quite impressive. Those who score very low on ordinary intelligence tests, the mentally deficient, do not solve the problems of their own lives very well, even if born with a silver spoon. They may, in fact, require institutional care just because they cannot look after themselves. Those who score very high on a standard test of general intelligence are far ahead, in life's struggles, of the mentally deficient, and even of average citizens (Terman and Oden, *1947*). The remarkable group of 1500 boys and girls with high IQ's that Terman has followed for 30 years must be considered good problem solvers by any reasonable standard that can be applied.

The use of intelligence tests in educational situations has shown that scores on these tests are related, by correlation coefficients in the neighborhood of .50, to academic success. In the employment situation, scores on tests of general intelligence predict success in solving problems encountered in professional, executive, and technical jobs. If the

problems are of an interpersonal nature, as in selling jobs, intelligence test scores are of only limited usefulness.

Many experimenters studying problem solving in the laboratory have given a test of general intelligence to their subjects. In practically all these cases those who solve the problems have higher intelligence test scores that those who fail. The magnitude of the correlation is usually not known because the reliability of the problem-solving scores is not known. If a wide variety of problems were used in sufficient number that a good measure of problem-solving success could be obtained, there is no doubt that such scores would correlate closely with scores on a conventional test of general intelligence. In Billings's (*1934*) study, described above, the eight tests of problem-solving ability showed a median correlation of .54 with a test of general intelligence.

GROUP FACTORS AND PROBLEM-SOLVING ABILITIES

Correlational analysis, especially factor analysis, has shown that it is possible to sort tests into several more or less separate groups. Correlations between tests within each group are relatively high, presumably because success on each test is based on a common ability or group factor. The most generally accepted group factors are the verbal, numerical, and spatial factors, so named because of the content or material of the tests in which they are found. The question arises, then, of the relation between each of these abilities and the ability to solve problems involving each kind of material.

The question can be clearly posed in relation to numerical material. It has long been known that ability to do simple arithmetic, e.g., addition, subtraction, multiplication, and division, is different from the ability to solve arithmetic problems. Standardized arithmetic tests used in high school, for example, frequently yield two scores, one called arithmetic computation, and the other called arithmetic problem solving or arithmetic reasoning. Correlations between these two scores are usually around .40 in high-school samples, or .50 when corrected for attenuation. Arithmetic reasoning is more closely correlated with general intelligence than arithmetic computation, though this statement does not mean much because tests of general intelligence include arithmetic reasoning items. Scores for arithmetic computation represent, in the terminology of Chapter 5, uncontrolled past learning,

while scores for arithmetic reasoning represent simple arithmetic computation plus a more complex problem-solving ability.

In the verbal domain there are no tests that correspond exactly to arithmetic computation. There is a limited number of arithmetic combinations, e.g., $3 + 5 = 8$, so everyone who goes to school has an opportunity to learn these combinations, and there is a well-defined population of combinations to be sampled for a test. But there are no comparable standard word combinations to which everyone is exposed. Tests of word fluency, simple word meanings, and common facts may be taken, however, as sampling uncontrolled past learning of simple verbal material. Tests of verbal problem solving consist of anagrams, crossword puzzles, abstract verbal relations, verbal analogies, interpretation of proverbs, and the like. Correlations between scores on simple verbal tests and verbal problem solving are approximately of the same magnitude as the corresponding correlations in the arithmetic domain. And of course the more complex tests have the higher correlations with tests of general intelligence.

In the domain of spatial abilities, problems are easy to develop. In fact many of the problems used by experimental psychologists have been spatial and mechanical puzzles. The space factor identified by factor analysis, though it may not be as clear-cut as originally supposed, seems to be not so much a sampling of past learning, like the verbal and numerical factors, as a measure of a perceptual ability. It is less closely related to the verbal and numerical abilities than these two are to each other. Thus the spatial factor makes less of a contribution to scores for general intelligence than the other two factors. Unfortunately, the relation between the space factor and success in solving geometrical and construction problems has not been thoroughly studied. What evidence there is (e.g., Harrell, *1940*) indicates that there is a definite relationship. Analysis of the behavior and comments of subjects solving spatial problems may help define the spatial factors (Barratt, *1953*).

In summary, a fairly simple picture of the relations between these simple group factors and success in solving problems can be outlined. Each of these three factors is more or less independent of the others when tested in its purest and simplest manifestation. Scores on each of these factors are related to solving problems involving these three kinds of material. In each area the more complex tests that involve some

problem solving are more closely related to each other and to general intelligence than tests that are pure and simple factor measures.

At least two reasons—not unrelated—may be advanced to account for the fact that complex problem-solving tests involving different materials are more closely related than simple tests of the same materials. First, the more complex tests depend more on general intellectual ability. Some additional ability is required for complex intellectual operations, and this additional ability is a general ability, e.g., an ability for abstract thinking or reasoning.

The second reason has frequently been offered by factor analysts when explaining unexpected results. As tasks become more complex, the abilities required for their execution become confounded. We have noted in Chapter 8 that the solution of verbal problems may be facilitated by numerical and spatial aids. Numerical problems often appear in verbal form, and verbal and spatial aids may be helpful. The solution of spatial and mechanical puzzles is likewise furthered by verbal formulation and numerical treatment. Hence it may not be possible to construct complex problems that are factorially simple.

Verbal, numerical, and spatial abilities are named for the materials or contents in the manipulation of which these abilities are manifested. Other abilities have been named for psychological processes or operations, such as the attention factor (see page 92) and the memory factor (see page 137). Among the important process abilities that have come in for some study recently are three of considerable significance for this book: reasoning, judgment, and creative thinking.

REASONING

This term reasoning has been used in many ways by philosophers, psychologists, and educators. As used in psychology today it usually refers to the solution of problems by logical operations. The notion of problem is included in the definition because the term reasoning refers not to all logical operations, only to those that solve problems. Adding two and two to get four is certainly logical, but the activity would not be called reasoning by most psychologists because it is not a problem. The first response is accepted. The notion of logical refers to the kind of operation most prominent in the problem-solving activity. Reasoning is contrasted with associative learning, as by Maier (see page 121), with recall, and with trial and error.

A Note on Methodology. In the terminology of preceding chap-

ters the thinker, in this case the reasoner, organizes the problem situation and the wanted solution as well as he can. Then he produces responses, in this case abstract patterns, and later judges their acceptability. In the special case of problem-solving activity called reasoning, the organization of the problem situation leads directly to production of abstract patterns. The search model is relatively adequate and takes the form generally of an inference: If thus and so, then ———. Organizing the particulars of the problem situation into a conceptual pattern may be called induction. Fitting a solution pattern to the search model thus organized may be called deduction.

Whatever terms are used, individual differences may be apparent at any point in the sequence of logical operations. The customary strategy is to devise tasks in which only one logical operation is difficult. Hence scores for success on these tasks are therefore measures of success on this operation. Consider the verbal analogy:

flower is to petal as table is to_____
floor, leg, chair, plate, bud

The first step here presumably is discovering a relationship between flower and petal. This could be called induction, but in general terms let us call it simply Step A. The next step is fitting table into this relationship with a gap for the other term. This could be called the construction of a search model, but let us call it simply Step B. Next the thinker tries out the given words to see which one best fits the search model. This could be called deduction, but let us call it Step C. This analysis may not be the best one, but any other defensible analysis would also consist of two or more steps, all of which are necessary for attainment of the solution. The scores on a test of such items therefore measure ability to perform all the steps.

If we wish to get measures of ability to perform Step C, it is necessary to construct items in which Step C is sufficiently difficult that some will fail at this point, while Steps A and B are so easy that everyone will be successful at these points. The ability to perform Step B can be tested by making Step B difficult and Steps A and C easy. And so on. The strategy is a general one which can be followed in testing for ability to perform any logical operation that can be identified, even though the performance of the tasks includes other logical operations. If the attempt to follow this strategy fails, it may be because the logical analysis on which it is based is incorrect or because the relative difficulty of the various operations is incorrectly estimated.

Research on individual differences in reasoning has proceeded according to this strategy in part but not exclusively. Tests have been constructed on the basis of this rationale—not always explicitly to be sure—then correlations have been computed to see if the expected relations appear. When the relationships that appear are not as expected, *ex post facto* analysis is often applied. The investigator reexamines a test and now sees that it was not really a measure of Step A, but of Step B or perhaps Step D. This *ex post facto* procedure is common in factor analysis—and in any kind of research where the important variables are not well identified in advance. Reasoning factors, or any others, identified by this retrospective procedure should be considered hypotheses subject to confirmation in further researches free from this logical weakness.

Research on individual differences in reasoning has been carried on principally by factor analysis and has been directed toward several important theoretical and practical questions.

1. Can reasoning ability or abilities be identified by factor analysis? If so, are such abilities limited to one kind of material or do they cut across verbal, numerical, and spatial material? What kinds of intellectual operations are involved? For example, is induction different from deduction in this respect?

2. How are these abilities related to other well-known abilities? The relation between reasoning and the general intellectual factor, for example, is an interesting controversial question. Years ago Burt (*1909*) used syllogisms and similar logical tasks in a battery of tests for school children and stated that the syllogism was the best single test of general intelligence. Spearman (*1927*) has long maintained that *g*, the general factor found in tests of intellectual functions, may be considered the eduction of relations and correlates.

3. How are these reasoning abilities related to success in academic and nonacademic activities? Do tests of these abilities have any practical predictive value not possessed by other well-known tests?

IDENTIFICATION OF REASONING FACTORS

When Thurstone (*1938*) applied his multiple-factor method to the analysis of the correlations between 57 tests taken by 250 college students, a factor called induction was identified. It was best tested by the number series test:

Find the rule in the series below and fill in the blanks.
10 8 11 — 12 — 13

Two other factors, not clearly outlined, were tentatively called restrictive thinking, or reasoning, and deduction. Arithmetic reasoning was most prominent in the former and syllogisms in the latter. A later factor analysis of 60 tests taken by 710 eighth-grade pupils turned up only the induction factor (Thurstone and Thurstone, *1941*).

Various batteries of tests of group factors or primary abilities, prepared for general use, have included tests of these types, yielding scores for induction and deduction, or just one score for reasoning. For example the Chicago Tests of Primary Mental Abilities, for ages 11 to 17 (Thurstone and Thurstone, *1943*), get reasoning scores from two tests, letter series and letter grouping.

Letter series presents the subject a series of letters in which he is to discover a rule. Then he can reason out what letter should come next, and choose that letter from printed alternatives.

What letter should come next?
a b m c d m e f m g h m — g h i j m n

Letter grouping asks the subject to mark the group of letters that is different in some way from the others.

In this example,
AABC ACAD AGFH AACG
the third group is to be marked because each of the other groups has two A's. Here is a harder item:
BDEF ILMN LNOP QSTU

Many reasoning tests have been presented in verbal form, e.g., syllogisms and arithmetical reasoning. Blakey (*1941*) raised the question, therefore, whether a reasoning factor could be found in nonverbal tests. The tests were printed on paper and the instructions were given verbally, but the items consisted, not of words, but of forms to be compared, circles to be marked according to a system, and other objects to be classified and treated according to abstract rules. The tests were given to 286 high-school subjects.

Five factors were extracted, some of which were perceptual factors, but the most important factor was a general one found in all tests, though the loadings of some tests were quite small. The best test of this factor was form reasoning, which required the subject to combine

three forms to equal another according to a rule. The test with the next highest loading was circle grouping, which required discovering a system for marking circles. And another test with a high loading was form series, analogous to the number and letter series illustrated above. Some of these tests, or tests like them, have been used in other factor analysis of reasoning, permitting the conclusion that nonverbal reasoning and verbal reasoning have much in common. The essence of the reasoning factor seems to be, not the material on content or appearance of the items, but the abstract nature of the operations performed on these items.

AAF Studies of Reasoning. The mission of the Army Air Forces program of test development and analysis was improvement of selection and classification of pilots, bombardiers, and navigators. The AAF Qualifying Examination (Davis, *1947*) was given to civilians and servicemen to select men who could graduate from flying training, especially pilot training, since pilots were in greatest demand in 1942.

Tests of reasoning and practical judgment were included in the examination and proved to contribute significantly to its validity. Factor analysis of the scores of 689 high-school boys, carried out by the principal-axis method, emphasized a fairly generalized reasoning ability, "probably closely akin to the principal mental function measured by ordinary intelligence tests." This was a component of all tests believed to measure judgment and reasoning, but was strongest in tests called arithmetical reasoning, figure analogies, pattern analogies, reasoning judgment, inferences in reasoning, syllogistic reasoning, and logical reasoning. A judgment component and other reasoning components of less importance were identified but not very clearly defined. (Good judgment as an ability will be discussed below.)

Printed classification tests were developed by the Army Air Forces (Guilford, *1947*) for the classification of Air Forces flying personnel. This program stressed the construction of a wide variety of tests, correlation with military criteria, factor analysis of intercorrelations by Thurstone's centroid method, and refinement in the direction of uniqueness.

Among the requirements frequently mentioned by flight instructors and supervisors were arithmetic reasoning, especially for navigators, ability to adjust to new situations, good judgment, foresight, and comprehension. It seemed likely that some of these characteristics of good airmen could be tested by reasoning tests.

A fairly difficult item from a test of arithmetic reasoning follows:

A plane traveled a certain distance from the base at an average rate of 225 miles per hour. Engine trouble forced it to return at an average rate of 150 miles per hour. It left at 11:35 A.M. and returned at 12:05 P.M. How far away from its base was the plane when it turned back? A. 30 miles. B. 45 miles. C. 50 miles. D. 75 miles. E. 90 miles.

Analysis showed that an arithmetic reasoning test of this type does not depend upon mathematical training as much as ordinary mathematical achievement tests. This test is one of the best for predicting success as navigator, but has little validity for pilots. Its highest factor loadings are in the numerical and the general reasoning factors. The number series test, illustrated above, likewise has its highest loadings in these two factors.

As a fairly typical example of a nonverbal, nonnumerical type of reasoning item, see Figures 11.1 and 11.2, taken from an AAF test of decoding. Items of this sort may, of course, be made quite difficult; instructions are lengthy. Factorially, these tests are loaded principally in reasoning and perceptual factors.

Other types of nonverbal, nonnumerical reasoning tests require classification of geometrical figures, discovery and use of figure analogies, completion of symbol patterns according to a rule, and the like. The items are visible but the relations between them are abstract and not obvious.

Factor analysis of the scores of 202 pilots on tests like the above, together with tests of known factorial composition, brought out a general reasoning factor common to practically all reasoning tests. It resembles Thurstone's restrictive reasoning factor (see page 399). No tests were pure tests of general reasoning; all were factorially complex. Some included the numerical factor, some the verbal factor, some the spatial factor, and so on. And narrower reasoning factors appearing only in a few tests were suggested, one of which resembled Thurstone's induction.

Other factor analyses (e.g., Holzinger and Harman, *1938*) have not confirmed the separation of inductive and deductive reasoning. Factor analysis by the bifactor method (Holzinger and Swineford, *1939*) of tests of reasoning, deduction, problem solving, and the like along with other kinds of tests failed to bring out a group factor of reasoning at all. These tests of thinking, on the contrary, had high loadings on the general factor, and Holzinger and Swineford conclude, therefore, that

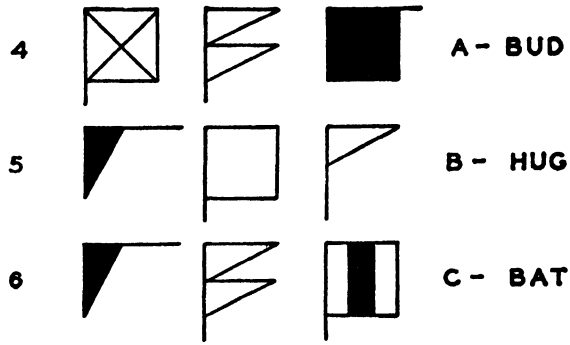


FIGURE 11.1. A nonverbal, nonnumerical reasoning problem from an AAF test of decoding (Guilford, *1947*). Instructions:

Note that the letter *b* appears twice, each time at the beginning of a word. Since the symbol which appears at the beginning of two code words is a black pennant . . . this pennant must represent *b*. The letter *u* also occurs twice in these words, both times in the middle of a word. Thus the code symbol for *u* is the double white pennant. . . . Since the first two symbols of item 6 stand for *b* and *u*, this item must be *bud*. The other two items are solved by noting that item 5 begins with *b* and is therefore *bat*, and that item 4 contains *u* as a middle letter and, therefore, must be *hug*. . . .

the general factor consists largely of reasoning ability. (Note also Davis's statement above.) This conclusion backs up Spearman's original claim for *g*, and it agrees with the common statement that ability to handle abstract relations is at the center of intelligence.

USC STUDY OF REASONING. Probably the most thorough factor analysis of reasoning to date is that carried out by Guilford and associates (Guilford, Comrey, Green, and Christensen, *1950;* Guilford, Green, and Christensen, *1951*) at the University of Southern California. The factor analyses summarized above led these investigators to several hypotheses about the nature of the general reasoning ability. It might be the ability to manipulate symbols; the ability to solve problems; the ability to define, formulate, or structure given problems; the ability to test hypotheses; or the ability to organize a sequence of logical operations.

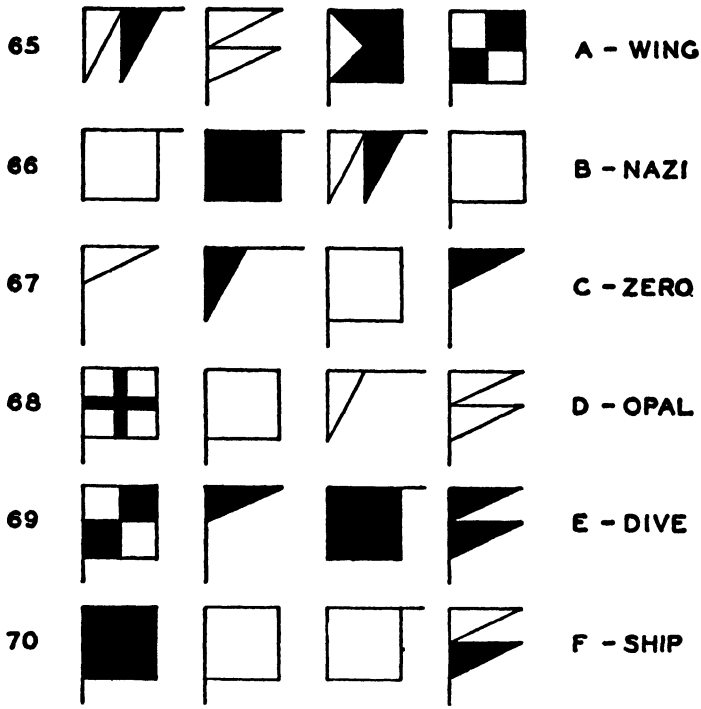


FIGURE 11.2. A more difficult nonverbal, nonnumerical reasoning problem. Instructions as for Figure 11.1. (Guilford, *1947*).

A second reasoning ability could be called induction. More specific and analytical hypotheses about this ability were that it is the ability to grasp a system of relationships in its totality, or that is an ability to see trends in a series of objects, or to identify a relationship in a variety of settings. Or, since the Gottschaldt test of hidden figures has been found to be related to reasoning, this reasoning ability might be a closure factor common to perceptual and symbolic material.

Another hypothetical reasoning ability could be the ability to classify objects, either a general classifying ability or a group of abilities for classifying different kinds of objects. And another reasoning ability could be deduction, as in the syllogisms test.

To check on these hypotheses 34 short tests were used, most of which were based on these hypotheses, while others tested well-known reference factors, such as number facility and verbal comprehension. As an illustration of the care with which these tests were constructed,

we may contrast Verbal Analogies I with Verbal Analogies II. The former was designed to test eduction of relations. Therefore the logical operation involved in finding the relation between the first pair of words was made difficult, while other logical operations were made relatively easy:

electricity: energy :: green :_____?
color red grass trees foliage

Analogies II was designed to test eduction of correlates. Therefore the relation between the first pair of words was obvious, while the relation between the second pair was relatively difficult:

dog : legs :: bank :_____?
checks vault government credit depositors

These 34 tests were given to two military samples, a total of 283 subjects, selected for special training on the basis of intelligence test and educational requirements. The correlations were factored by the centroid method, with rotation to orthogonal axes.

The most important factor identified was the general reasoning factor. This is a rather broad factor in which reasoning about arithmetic problems as usual was the chief defining variable. Reasoning about spatial or perceptual material was also well represented. Reasoning of a deductive nature, as in syllogisms, was absent—by one analysis at least. These results led the investigators to a new hypothesis about the nature of this general reasoning factor:

> Most of the tests that appear on this factor involve the manipulation of abstract symbols which have little realistic meaning. As problems get harder, trial-and-error manipulation may reasonably be assumed to become more important. This factor, then, may represent a less meaningful manipulation of symbols in a trial-and-error fashion. This would explain why difficult items in almost any type of test introduces some general-reasoning variance. This conception is very close to Heidbreder's idea of "spectator behavior" [See page 238]. It may also be akin to Tolman's concept of VTE (vicarious trial and error). Effective trial-and-error approach would presumably be most dependent upon the speed with which new solutions are tried and on the ease with which failing solutions are rejected. . . .
>
> Thus, we have not solved the nature of reasoning I, but we have a number of new hypotheses to be investigated. It is definitely a factor that creeps into quite a variety of tests and hence must have some rather broad or pervasive properties.

A second reasoning factor was identified in a variety of tests, but chiefly in tests of deductive reasoning, inferences, syllogisms, and false premises. The investigators' interpretation of this factor follows, in part:

> Presumably the chief process in deductive reasoning is the drawing of inferences or conclusions. But note that the tests that have been used to define the so-called deductive factor have been of the multiple-choice or true-false form. The examinee does not have to draw his own conclusion. Conclusions are given to him and what he must do is to decide which one is correct. Deciding about the correctness of a conclusion is a different process than producing the conclusion. It is an act of judgment or evaluation. The criterion of evaluation in these tests is that of logical necessity. We might describe this factor as a sensitivity to logical necessity. The name we have chosen, "logical reasoning," is of broader connotation to allow for other and possibly more exact descriptions. It remains to be seen whether deduction tests in the form of completion items will give rise to an additional factor separate from this one.

Other factors, of lesser importance, were called eduction of perceptual relations, eduction of conceptual relations, eduction of conceptual patterns, eduction of correlates, and symbol substitution.

Looking back at the hypotheses with which this investigation began, Guilford and his colleagues believe that the general reasoning factor has been substantiated, as in Thurstone's original work and in the AAF studies, though its nature is not well understood. It is not as broad as all problem solving, nor as narrow as arithmetic problem solving alone. The emphasis on trial and error in the first quotation above would cause some critics to question whether it is reasoning at all. It could, of course, be the general factor, or *g*, of the British factor analysts (see Vernon, *1950*), but the investigation at hand was not directed toward this question.

No general induction factor was identified, nor any classification factor. Factors that included tests in these areas were called eduction factors, after Spearman. Manipulation of symbols was not a broad factor but a rather narrow one. No factor of closure or analysis of forms was identified, but hidden figures came out on the factor called eduction of correlates.

Perhaps the most interesting outcome was that tests of syllogisms and the like, which have been considered typical of reasoning, are not prominent in the general reasoning factor. Instead they form a factor

by themselves which is not called deduction, for reasons given above, but logical reasoning.

UNC Study of Reasoning. Adkins and Lyerly (*1952*) at the University of North Carolina used a representative sample of Army enlisted men for an analysis of a variety of reasoning tests. In addition to the well-known factors, several appeared that could be described as reasoning factors.

One called perception of abstract similarities was loaded principally with classification and analogies tests.

One called hypothesis verification included Raven's Progressive Matrices tests and others that have previously been listed under general reasoning. The present name was preferred because some subjects approached the problems by checking hypotheses suggested by the alternative answers.

Another factor was called ability to generalize, or form concepts, and there was also a small deduction factor.

These factors are not as independent as those found in previous studies by the centroid method, presumably because the sample was not so highly selected.

An Attempt at Integration. Factor analysis aims to simplify the description of abilities. But as factor analyses are carried out with different tests, on different samples of subjects, by different analysts, the outcome may appear as confusing as the input. In an attempt to improve this situation Guilford, Green, Christensen, Hertzka, and Kettner (*1952*) at the University of Southern California have analyzed a comprehensive battery of tests, including many that have been used in previous studies of reasoning summarized in this chapter.

The results of this analysis and three previous analyses have been aligned in tabular form by Guilford and his colleagues. Table 11.1 is the pertinent portion of this attempt at integration. The question marks indicate that the identification is tentative. The picture presented by this alignment shows that it is possible to achieve some order in this domain. The fact that three of the four studies were directed by Guilford should be considered in evaluating the amount of agreement.

The AAF study, the first USC study, and the UNC study are briefly described in the text. The second USC study (Guilford, Green, Christensen, Hertzka, and Kettner, *1952*), from which this table was taken, attempted to integrate the results of the three earlier studies.

TABLE 11.1 Parallel Factors in Four Studies of Reasoning

AAF	USC_1	UNC	USC_2
Verbal comprehension	Verbal comprehension	Verbal relations	Verbal comprehension
Numerical facility	Number facility	Number	Numerical facility
Perceptual speed	Perceptual speed	Perceptual speed	Perceptual speed
Visualization	Visualization	Space 2	Visualization
Spatial relations	Spatial orientation	Space 1	Spatial orientation
General reasoning	General reasoning	Hypothesis verification?	General reasoning
	Logical reasoning	Deduction?	Logical reasoning
Reasoning III?	Eduction of correlates	Concept formation?	Eduction of correlates
	Eduction of conceptual relations	Perception of abstract similarities?	Eduction of conceptual relations
Reasoning II?	Eduction of perceptual relations		Eduction of perceptual relations
Judgment			Judgment

CRITICAL THINKING. For many years Watson and Glaser, among others, have attempted to test verbal reasoning, or critical thinking, in relation to social problems. The Watson-Glaser Critical Thinking Appraisal is the latest in a series of efforts in this direction (Watson, *1925;* Glaser, *1941;* Watson and Glaser, *1949*). "Its items are mostly of a realistic type, involving problems, statements, arguments, and interpretation of data similar to those which a citizen in a democracy might encounter in his daily life as he works, reads the newspaper, hears speeches, participates in discussions on various issues, et cetera." In present form the Appraisal has five tests, called Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. These five tests are more or less independent; that is, the ten intercorrelations, when corrected for attenuation, range between .40 and .74, and half of them below .50.

Intercorrelations obtained from a large sample of college students show that the inference test, within the limits of its low reliability, is closely related to verbal ability in general, more closely in fact than the total score. The tests of recognition of assumptions, deduction, and interpretation cluster together and are relatively independent of the verbal factor. Hence it is likely that these three tests are loaded on the general reasoning factor.

Unfortunately the scores on these five tests have low reliabilities. The reliability of the total score for college students was .83. Whatever the total score may measure, it is relatively independent of conventional intellectual abilities. In a college sample, for example, the total score correlated only .41 with general intelligence and only .38 with reading ability. The chief distinction of this instrument, in comparison with other reasoning tests, is its breadth, since it includes arguments about such diverse issues as socialism, the size of heavyweight boxers, radiation sickness, divorce, race differences in intelligence, and women drivers.

REASONING AND OTHER FACTORS

We have seen that the general reasoning factor transcends the content of the test. Tests that are loaded on reasoning are usually loaded to some extent on the verbal, the numerical, or the spatial factor also, but certainly general reasoning is not highly related to any of these three content factors. Guilford's narrower logical reasoning or deduction factor has some special relation to the verbal factor, however, for several tests appear on both. Two studies have been specifically directed toward the relation between reasoning and closure, and some evidence on the relation between reasoning and general intelligence can be gleaned from several studies.

REASONING AND CLOSURE. As noted earlier (p. 111), closure factors were tentatively identified by Thurstone (*1944*) in his factorial study of perception: flexibility of closure, speed of closure, and strength of configuration. The flexibility-of-closure factor was described as "the ability to manipulate several more or less irrelevant or conflicting gestalts or configurations." Subsequent factor analyses have added some support to the identification of such factors (Meili, *1946;* M. D. Vernon, *1947;* Rimoldi, *1948*).

Thurstone's study of perception indicated that closure was associated with the reasoning factor called induction. Rimoldi's results are similar.

Botzum (*1951*) undertook to investigate this association more thoroughly. He gave 22 reasoning tests and 14 closure tests, along with 10 tests of well-known reference factors, to 237 college students and factored the correlations by the centroid method. Rotation was to oblique axes, according to the principles of simple structure.

In this oblique analysis a broad reasoning factor emerged, which was called induction. It is not the general reasoning factor of the AAF analysis, according to Botzum, but it has some tests in common with the general reasoning factor in the USC study, described above, by Guilford and associates. A second reasoning factor was called deduction because of the prominence of a test of false premises and one of syllogisms. It closely resembles Guilford's logical reasoning factor.

One factor was identified as flexibility of closure by the presence of a copying test and the Gottschaldt hidden figures test. A second closure factor included the Street Gestalt test, a test of backward writing, one of mutilated words, and the like. It was therefore called speed of closure. "In all of the tests with loadings on this factor, there is an unstructured field, in which some reorganization must occur."

In this study the factors are not independent. Induction correlated .498 with deduction, .427 with flexibility of closure, and .503 with the space factor. Flexibility of closure correlated .427 with induction, .262 with deduction, and .672 with the space factor. Thus it was possible to identify a second-order factor made up of these four first-order factors.

A similar factor analysis by Pemberton (*1952*) also disclosed the flexibility-of-closure and the speed-of-closure factors. Another factor was called verbal closure because its highest loadings were in anagrams, scrambled words, and incomplete words. As in Botzum's study, the reasoning tests were associated with flexibility of closure. Also in the UNC study by Adkins and Lyerly, mentioned above, factors called flexibility of closure and speed of closure were identified by similar tests. When the factors resulting from this oblique solution were factored (Matin and Adkins), the reasoning factor called perception of abstract similarities and the factor called flexibility of perceptual closure defined a second-order factor called flexibility in analysis.

These three analyses, therefore, add some support to the arguments of Binet and the Gestalt psychologists (see Chapter 4) for a connection between perceptual reorganization and cognitive reorganization.

Reasoning and General Intelligence. Correlations between tests of abilities are always positive, and some are higher than others. If we

wish to sort a large number of tests into small groups of related tests, we can (1) first take out the general factor in all tests, then look for small group factors; or (2) we can look for the group factors, and then compute a second-order general factor. British factorists by and large have preferred the first alternative while American factorists have preferred the second. Mathematically the two solutions are interchangeable, hence a preference for either one flows from motives ulterior to the mathematical relationships, e.g., purpose of the analysis, and assumptions about the universe.

The question of the relation between reasoning and general intelligence thus becomes not so much a question of fact as one of alternative ways of viewing the facts. Some facts for consideration are printed in Table 11.2, taken from the manual of instructions of the Chicago Tests of Primary Mental Abilities (*1943*). These are based on the

TABLE 11.2. Correlations Between Thurstone's Reasoning Factor and Other Primary Factors [a]

	N	W	V	S	M	R
Reasoning	.540	.480	.548	.386	.389	
General factor	.603	.686	.676	.339	.474	.843

[a] From Thurstone and Thurstone (*1943*).

scores of 1000 Chicago students between the ages of 10 and 18. Correlations between the reasoning factor and the number, word fluency, verbal relations, space, and memory factors are shown. The reasoning scores come from the two tests illustrated on page 399, letter series and letter groups, which are not pure measures of the reasoning factor. The general factor is found by factoring the table of correlations between the primary factors. It is apparent that the closest associations of the reasoning factor are with the verbal and numerical factors. More important, it is the reasoning factor that has the highest correlation with the general factor. If by general intelligence we mean a second-order factor obtained by factoring primary abilities, reasoning is closely related to this general intelligence.

But by general intelligence we may mean the score on a conventional test of general intelligence. Tredick has reported (see Goodman, *1944*) correlations between primary abilities and several standard tests, including tests of general intelligence devised by Otis and by Pressey, for 113 female college freshmen. (In this battery induction and deduction are scored separately.) The highest correlations with the Pressey test are .76 for the verbal factor, .52 for induction, and .61 for

deduction. The corresponding correlations for the Otis test are .68, .60, and .61. Correlations with other factors are low. It is evident that scores on reasoning tests are related to scores on ordinary tests of general intelligence, but the relationship between the verbal factor and general intelligence is somewhat higher.

Recent data obtained with the SRA Primary Mental Abilities battery shed further light on this question (Thurstone and Thurstone, *1949*). In this battery reasoning scores come from only one test, a six-minute test of letter series. The verbal factor is tested by a four-minute vocabulary test. In a sample of 92 tenth-grade pupils Otis IQ scores correlated .48 with the verbal factor, .30 with the reasoning factor, and .09 with the number factor. The multiple correlation was .71. In a sample of 142 high-school students, Kuhlman-Anderson IQ correlated .52 with the verbal factor and .22 with the reasoning factor. The multiple correlation was .63. This sample had taken the intelligence test in the eighth grade. And a study by Jacobs (*1948*) with the American Council on Education Psychological Examination yielded correlations for 97 tenth-grade pupils of .67 with the verbal factor and .58 with reasoning.

It is clear, then, that the communality between scores on tests of general intelligence and scores on primary abilities is accounted for chiefly by the verbal and reasoning factors, with the verbal factor somewhat more important. Or, otherwise stated, most tests of general intelligence test primarily the verbal factor, secondarily the reasoning factor.

A similar picture of the relation between reasoning and general intelligence comes from a factor analysis of the Stanford-Binet Test by Jones (*1949*). He used the scores of the standardization sample of the 1937 revision, factoring the correlations by the centroid method. The verbal factor was the most prominent, as usual; but he identified in addition, at most age levels, poorly defined reasoning factors, represented by such tests as plan of search, ingenuity, verbal absurdities, similarities and differences, and comprehension. If these are reasoning factors, it is important to note that they rank next to the verbal factor in contribution to the magnitude of the IQ. They are not clearly separated from the verbal factor; several tests appear on both.

British factor analysts usually prefer to extract the general factor first, then to look for group factors. Much of the variance in the scores that the multiple-factor method attributes to reasoning is taken out with

the general factor, *g*. For example, when Vernon (*1950*) reanalyzed the data for the 689 high-school boys (page 400), he found that "a *g* (with some 22 percent variance) + group factors for verbal, mechanical and spatial-perceptual tests (totaling some 12 percent) gave almost as good a fit as Guilford's six multiple factors—V, mechanical, visualization, general reasoning, reasoning II, and judgment." Vernon admits that "a small reasoning or logic group factor could be isolated from specialized tests, but it would be unlikely to add anything to measures of *g*, *v*, and *n* in the prediction of the reasoning ability desirable among secondary pupils or college students." There is in fact some evidence for one or two reasoning factors after the extraction of *g* from a battery of tests given to French high-school boys (Bernyer, *1946–1947*).

Whether to extract *g* first and then look for group factors or to avoid *g* and look directly for multiple factors is largely a matter of convenience and prior commitments. When the latter procedure is followed, one or more reasoning factors can often be identified, though not clearly defined.

REASONING TESTS AS PREDICTORS

The question of validity revolves around the use of reasoning tests to predict achievement. A number of investigations of correlations between the Chicago Tests of Primary Mental Abilities and college grades have been reported by Goodman (*1944*). In general, reasoning has its highest correlations with grades in science and mathematics courses. For most courses reasoning again plays second fiddle to the verbal factor. Some data given by Goodman permit a comparison of correlations of primary abilities with grades for the first semester and with grades for the first two years. As in other similar studies the correlations, including the correlations between inductive reasoning and grades, show a general drop with the passage of time. Correlations between deductive reasoning and grades, however, increase for all courses. This reversal of the usual finding suggests that as the courses become more difficult, the contribution of deductive reasoning increases. Incidentally it adds some encouragement to the attempt to separate inductive from deductive reasoning.

When scores on the SRA Primary Mental Abilities Test were compared with total scores on the Iowa Tests of Educational Development (Thurstone and Thurstone, *1949*), the verbal factor (a four-minute test of vocabulary) had the highest predictive power, as represented

by a correlation of .70, and the reasoning factor (a six-minute test of letter series) was next, with a correlation of .46. The power of primary abilities to predict high-school achievement must come largely from the verbal factor, for the multiple correlation based on the verbal factor and the reasoning factor was .72, scarcely more than the correlation with the verbal factor alone.

The subtest of the Iowa battery with which the reasoning factor correlated highest was one called Correctness and Appropriateness of Expression. The reasoning factor correlated .52 with this test, the verbal factor .55. With a test called Ability to Do Quantitative Thinking, the reasoning factor correlated .42, the verbal factor .43, and the number factor .27.

We are led to the conclusion that the reasoning factor of Primary Mental Abilities has some validity as a predictor of success in high-school and college courses. It is not so useful, however, as the verbal factor. Its correlations with achievement are not specialized; like the verbal factor it has respectable correlations with success in almost all courses. If a test of general intelligence is used first, or if the general factor, *g*, is removed first, would a residual reasoning factor contribute anything to prediction of college grades? Apparently this question has not been seriously studied.

Validities of the reasoning tests developed in the AAF psychology program (Guilford, *1947*) were computed for large samples of pilot and navigator trainees. The criterion in each case was graduation from a regular military training course. Tests of arithmetic reasoning were quite valid against the navigator criterion, above .30 in several samples, but no reasoning test was valid for pilot trainees. Estimated validity of a general reasoning factor, based on tests of arithmetic reasoning and reading comprehension, for the navigator criterion was .38 in one study. Attempts to improve reasoning tests by purifying them of numerical and verbal content were moderately successful. The decoding test illustrated on page 402, for example, had validities of .20 to .24 for several groups of navigation students and .01 to .14 for several groups of pilots in primary training.

One might raise the question whether reasoning tests are superfluous for selection purposes. If the variance of these tests is accounted for by verbal, numerical, spatial, and other more clearly defined factors, scores on reasoning tests could be estimated from a battery of the other tests, and any predictions made from reasoning tests could

be made from appropriate combinations of the others. But the communality of the reasoning tests in the AAF program is not entirely accounted for by the other factors. Although they are not factorially pure, the reasoning tests have some factorial content and some predictive power not found in simpler tests. This finding is of some consequence in itself to those interested in individual differences in thinking, even though it did not advance the AAF selection program.

The Miller Analogies Test (Miller, *1947*), widely used in the selection of students for graduate training in psychology, is of the familiar analogies form and undoubtedly has much in common with other reasoning tests. Like other reasoning tests, it has high correlations with general intelligence (.82 with the Ohio State Psychological Examination) and with verbal ability (.80 with the verbal factor of the Graduate Record Examination). It has demonstrated surprisingly good predictive power. Correlations in the .70's with scores for educational achievement in various fields have been reported (Miller, *1947*). For example, in a study of 51 Ph. D.'s in psychology from the University of Minnesota, scores on the Miller Analogies Test correlated .51 with faculty ratings of intellectual competence, .27 with ratings of research competence, and .17 with ratings of service competence (Watters and Paterson, *1953*). Hence, even in this highly selected group, all of whom have passed a series of academic hurdles, the Miller scores predict intellectual competence. As one would expect, the ability tested by this test is not so closely related to research competence and service competence. In another large-scale study (Kelly and Fiske, *1951*) the Miller was the best single predictor of success in training for clinical psychology. Its validity was higher than that of the reasoning test of Primary Mental Abilities, presumably because it is longer and broader.

The usefulness of a reasoning test for prediction of achievement in any particular field of knowledge can be sharpened by phrasing the items to include that particular field. The Minnesota Psycho-Analogies Test, for example (Levine, *1950*), is designed to predict achievement in psychology and therefore includes psychological material in the analogies.

Orchestra : Violinist :: Test : _____?
Battery Item analysis Item Validity

By this strategy the test attains validity coefficients of .51 to .74 in predicting success in advanced psychology courses.

From the results at hand it is possible to describe the role of reasoning tests in prediction. To predict success in one course of study, e.g., in chemistry, celestial navigation, or in one occupation, e.g., clinical psychology, a broad omnibus test could be developed that would be superior to a reasoning test and superior to any other primary ability test. A good omnibus test would, of course, include some reasoning items; most good intelligence tests do. To predict success in several different courses or occupations, different tests of primary abilities are valuable, one of which could be a reasoning test. But reasoning tests overlap the other tests. The contribution of reasoning to a battery of tests is something like the contribution of tests of other primary abilities, but not so independent, and something like the contribution of a test of general intelligence, but not quite so general.

THOUGHTFUL TESTS OF EDUCATIONAL ACHIEVEMENT

No one believes that conventional tests of school achievement are as good as they should be. This dissatisfaction has encouraged development of achievement tests that place less emphasis on factual memory and more on critical thinking or reasoning. Tests of arithmetic reasoning, for example, are well known and known to be different from tests of arithmetic computation. Similar tests in other areas are being developed. Those who have the responsibility of testing achievement in physics, biology, psychology, history, or any other branch of knowledge have no excuse for limiting their examination to the more conventional types of items. Since achievement tests are often the criterion against which aptitude tests are validated, as in the investigations discussed above, obviously the validity of an aptitude test depends in part on the nature of the achievement test.

Some interesting tests that could be included under the general heading of reasoning were constructed by a staff of experts for the Progressive Education Association (Smith and Tyler, *1942*). They wished to have tests that would evaluate the results of new high-school courses aimed at improving thinking rather than accumulating information. Teachers were concerned about the ability of their pupils to make inferences, to apply principles, to argue fairly, and about tendencies to overgeneralize, or to be unduly cautious.

The items devised for this purpose resemble the items on reading tests. The items on one test, for example, present a few sentences, and perhaps a graph or a table, followed by pertinent statements to be

judged, in relation to the material at hand, as *true, false, probably true, probably false,* or as based on *insufficient data.* Scores are obtained in this way for accuracy in general and accuracy with specific types of statement; for a tendency to go beyond the data, e.g., to accept as true statements that are only probably true; for caution as indicated by marking a true statement *probably true;* and so on. Reliability coefficients for most of these scores are in the .80's.

In an attempt to get at the meaning of these scores, the investigators compared them with scores obtained from comparable essay tests. The essay tests presented statements to the subjects and let them draw their own conclusions. Correlations between scores on the essay tests and corresponding scores on the objective tests for 200 high-school students were above .80 for general accuracy when corrected for unreliability in the criterion. For the other scores, however, the correlations were rather low. Hence the comparison with the essay tests does not clarify the meaning of the scores on the objective tests.

When these different scores for 284 students are intercorrelated, an interesting correlation matrix emerges. The different accuracy scores are positively correlated, of course; there may be a general reasoning factor underlying success in all of them. But the caution score seems to introduce something else, perhaps a personality factor, that is not closely related to the other scores, and that seems to be related to accuracy on the *probably true, probably false,* and *insufficient data* items. The caution score obtained in this way may be related to confidence in a judgment, as discussed in the preceding chapter.

The evaluation staff also constructed tests of application of principles and of logical reasoning. Items on the test of application of principles describe a problem and give alternative answers, classified under these headings: correct principles, false facts and principles, irrelevant reasons, false analogies, popular misconceptions, the citing of unreliable authorities, ridicule, assuming the conclusion, and teleology. The reliabilities of scores for some of these subtests were fairly high when they were tried out; others were low. The same is true for the subtests of logical reasoning.

It must be admitted that we do not know just what these fascinating tests are measuring, with the exception of the most general score on each one, which is probably the reasoning factor found in the factor analyses described above. For the purpose of testing school achievement, these tests have sufficient face validity; but for present purposes

it would be desirable if scores from these tests could be correlated with scores from other tests of reasoning, from tests of difficult reading, intelligence tests, and above all with other measures of achievement. Factor analysis might clarify the relationships between scores and identify the abilities involved.

Some understanding of the contribution of reasoning to achievement in a college course in psychology may be gleaned from a recent study (Johnson and Smith, *1953*) using three types of test item: vocabulary, facts, and reasoning.

A vocabulary item:

SET: (1) preparation, (2) reaction, (3) pathway, (4) choice, (5) fixation.

A factual item:

Raising body temperature by electric current is a specific treatment for (1) psychosomatic illness, (2) psychoneurosis, (3) brain syphilis, (4) manic-depressive psychosis, (5) paranoid schizophrenia.

Two reasoning items:

The statement "Perception is an active process" means about the same as: (1) None are so blind as those who won't see. (2) See a pin, pick it up: all the day you'll have good luck. (3) "I see," said the blind man. "You're a liar," said the deaf man. (4) Look to the ant, thou sluggard. (5) Believe only half of what you see and nothing of what you hear.

Under which of these conditions will the sex drive in rats be stronger than hunger or thirst? (1) Place animal in obstruction box when young. (2) Deprive animal of food, water, and sex for two days. (3) Deprive animal of food and water for one day. (4) Keep male and female together constantly. (5) Give plenty of food and water.

Thirty items of each type were included in a test, given at the beginning of the course, and equivalent items were given at the end. A case study test and a psychology reading test were also used, both initially and finally.

The results obtained from 162 students before and after a beginning course in psychology showed that there is more overlap between initial and final scores on the reasoning test than on the tests of vocabulary and facts. In other words, a college course in psychology adds more to the students' vocabulary and factual knowledge than to their ability to use these in reasoning about psychological problems. Intercorrelations

between pairs of tests showed that the reasoning test had the most in common with the others. That is, psychological reasoning is the most general factor in psychological achievement. These correlations were higher, in general, at the beginning of the course than at the end. Apparently tests of vocabulary and facts were more or less reasoning tests at the beginning. Scores were quite low. At the end, after the students had acquired some information, less reasoning was required for these tests. In agreement with this inference is the finding (Sanua, *1952*) that scores for vocabulary and facts were more closely related to general intelligence at the beginning of the course than at the end. General intelligence was more closely related to reasoning scores, as would be expected, than to scores on any of the other tests.

An incidental inference can be drawn from these results to emphasize the importance of the criterion of achievement in validity studies. If the final achievement examination is made up only of vocabulary and factual items, as many are, scores on reasoning tests will not predict achievement very well. The validity of reasoning tests as predictors of achievement will improve as achievement tests improve. For example, when achievement in college physics is measured by the problem-solving type of final examination, the reasoning test of Primary Mental Abilities is a slightly better predictor than the verbal test (Dunning, *1949*).

It is true that reliable tests of the more thoughtful type are difficult to construct. Reliability coefficients are lower than for tests of facts and vocabulary constructed with equal care. But helpful examples and hints are available in such sources as the *45th Yearbook* of the National Society for the Study of Education, called *The Measurement of Understanding* (*1946*), as well as in the researches cited above.

The objection is sometimes raised that achievement tests requiring the student to make inferences, apply principles, and solve problems resemble intelligence tests more than they resemble the old-fashioned tests of facts and terms. Why not test achievement by giving an intelligence test and eliminate the nuisance of teaching a course in the subject? It is true that achievement tests of the thoughtful type correlate more highly with intelligence than the more conventional tests. But the objection puts the cart before the horse. If a good achievement test depends more on intelligence than a poor one, that relationship is not a weakness of the achievement test. It merely indicates one of the fundamentals of good achievement.

But the best reply to this objection is a more technical one. There is a real danger that tests of the problem-solving type may depend so much on reasoning that superior thinkers would attain maximum scores without studying the subject. The answer to this objection is to construct test items that require reasoning about the subject of the course. The problems should call for a knowledge of facts and reasoning about these facts. A kind of validation can be secured if the items are given to the students before and after a course of study. Items that are passed before the course as frequently as after the course are not valid measures of achievement in that course.

Another practical suggestion may be made at this point. When grades are given for educational achievement, one may wish to weight scores for vocabulary, facts, and reasoning equally. But when these scores are used to predict success in advanced courses, it may turn out that the reasoning scores merit extra weight.

READING AND REASONING

Reading a difficult paragraph and trying to comprehend its meaning involves almost all the intellectual operations discussed in this book. Years ago Thorndike (*1918*) analyzed errors made in reading and described the process of reading as follows:

> It consists in selecting the right elements of the situation and putting them together in the right relations, and also with the right amount of weight or influence or force for each. The mind is assailed as it were by every word in the paragraph. It must select, repress, soften, emphasize, correlate and organize, all under the influence of the right mental set or purpose or demand. . . . Reading an explanatory or argumentative paragraph . . . involves the same sort of organization and analytic action of ideas as occur in thinking of supposedly higher sorts.

Following this early suggestion by Thorndike, several attempts have been made to separate ordinary reading from some sort of reasoning process, e.g., from drawing inferences (Feder, *1938*), and from reference reading aimed at selecting answers to problems (Gans, *1940*). Langsam (*1941*) gave 21 tests, mostly various kinds of reading tests, to 100 college women and subjected the results to factor analysis. Five factors were extracted, one of which was called seeing relationships, but only the first or general reading factor was of any importance. Davis (*1944*) devised nine tests that appeared a priori to measure different components of the reading task and gave them to 421 college

freshmen. When he factored the correlations, by the principal-axes method, a large word knowledge or vocabulary factor appeared. Some variance was also attributable to a second component, called reasoning in reading, found in tests of drawing inferences. As usual in factor analysis the separation of these two factors is not completely satisfactory (Thurstone, *1946;* Davis, *1946*). Factor analysis of the AAF Qualifying Examination (Davis, *1947*) likewise disclosed a small factor called reasoning in reading (see page 400). The following item illustrates what is meant by this term:

One of the most beautiful military replies I've ever heard of was given in India by a captain who had lost a steam roller. The government sent him several forms to be filled out before it could be replaced. On one form was the question: "Reason for loss?" The captain filled in the words: "Eaten by white ants." He never heard another word about it, but in due course of time his replacement arrived. It is most probable that the captain

A. did not really know what happened to his steam roller.
B. told the truth about the steam roller.
C. was disgusted at having to fill out so many forms.
D. did not dare tell what had really happened to the steam roller.
E. did not care whether his steam roller was replaced.

The reading process ordinarily is performed rapidly and globally, calling on many abilities simultaneously. Apparently it is difficult to write items that test different reading abilities separately, as factor analysis requires. The contribution of various hypothetical abilities or factors to the reading task can be estimated, however, if an ordinary reading test is included in the battery of tests from which the various factors are identified. When this was done in a factor analysis of AAF classification tests (Guilford, *1947*), the reading tests had their principal variance on the verbal factor, as expected, but some small variance was contributed also by the general reasoning factor.

The study by Johnson and Smith (*1953*), mentioned above, also produced some evidence on the relation between reading and reasoning. At the beginning of a course in psychology a difficult psychological reading test correlated .33 with a psychological reasoning test and .38 with a psychological vocabulary test. At the end of the course the corresponding correlations with reading were .69 and .14. Apparently reading depends on reasoning and vocabulary to about the same degree before the subject matter is studied. After the course, when all students have fair vocabularies, the influence of individual differences in vocab-

ulary is lower, and this difficult reading test becomes principally a reasoning test.

We must conclude, from the evidence at hand, that performance on ordinary reading tests depends largely on verbal ability, as measured by vocabulary tests, but that reasoning ability makes at least a small contribution to this performance. When special reading tests are used, the contribution of reasoning may be larger. Furthermore some tests, called critical thinking, practical judgment, and work planning, to be described below, resemble difficult reading tests and probably require reasoning in larger amounts.

GOOD JUDGMENT AS AN ABILITY

When the abilities that distinguish competent people are under discussion, we hear much about good judgment. It is commonly listed as a desirable characteristic of executives, aircraft pilots, quarterbacks, and parents. But this trait of competent people has not been well defined. Good judgment may be merely a pleonasm for good. The systematic treatment of judgment outlined in previous chapters suggests several abilities that could characterize good judges.

Good judgment could refer to the ability to make accurate judgments of simple material, such as distances and heights. Presumably a competent outfielder has to be a good judge of the trajectory of a batted ball and a competent flyer has to be a good judge of his distance above the ground when landing. Such abilities are probably quite specific to the type of material being judged. In the AAF aviation psychology program (Guilford, *1947*), tests requiring estimations of distance were developed for group administration and from these a length estimation factor was identified. It proved to be related more closely to the perceptual speed factor than to general judgment factor to be described below.

More often the term good judgment refers to the ability to make good judgments of complex stimulus material, several aspects of which must be taken into account simultaneously. From previous discussion of the conditions of good judgment, it would appear that the important abilities required are (1) abstracting the pertinent data from the complex situation, (2) adopting and maintaining a set for these pertinent data, (3) attending to several data simultaneously, (4) weighting each appropriately, and (5) integrating all this information somehow so that it is related to one of the response alternatives rather than the

others. Presumably such judgment requires (6) delaying or inhibiting response, i.e., carefulness, caution, or deliberateness. Errors could be made in any of these aspects of complex judgment and some people could be better than others in any or all aspects.

But it does little good to identify and assign names to hypothetical abilities unless tests of these can be developed or at least imagined. We must raise the same questions raised in connection with reasoning abilities: the nature of good judgment as an ability, relations with other abilities, and relations to practical success. Some progress toward answers to these questions has been made in military and industrial situations.

IDENTIFICATION OF JUDGMENT FACTORS

The ultimate mission of the AAF program (Guilford, *1947*) of test development and analysis was, of course, improvement of the selection of pilots, bombardiers, and navigators. But there is a lesson here for the psychology of judgment in general which can best be appreciated by following through the development of judgment tests.

AAF Tests of Good Judgment. Poor judgment was one of the reasons most frequently mentioned by instructors for eliminating cadets from flying training. One reason for this apparently was the common practice of flight instructors of placing errors of perception, visualization, and reasoning all under the broad category of poor judgment. Therefore aviation psychologists asked flight instructors what they meant by good judgment. From their comments some typical definitions were written:

1. The ability to react immediately and appropriately to stimuli with which an individual is unacquainted.
2. "Headwork" or the ability to react correctly without deliberation, or the ability to fly without confusion in traffic and under unusual circumstances.
3. Knowledge, plus speed of reaction, plus freedom from emotional confusion.
4. The ability to react appropriately in a surprise situation.
5. Ability to grasp the situation as a whole, not being absorbed with minor details.

Tests constructed for the analysis of individual differences in good judgment include items of the type illustrated below, taken from the first form of the AAF Qualifying Examination.

A pilot has to make a forced landing near a mountain cabin. He finds that the nearest phone is at an isolated fire ranger's cabin 14 miles across the mountains to the north. It is winter. He sets out on foot for the ranger's cabin at 6 A.M., carrying food for only one meal. At 10 A.M., having met no one, he comes to three branches of the trail, all unmarked. His most practical decision would be to:

A. Follow the trail which appears to lead in the right direction until he reaches the cabin or the end of the trail.
B. Turn back immediately to his starting point.
C. Leave the trail and go due north by compass.
D. Walk along the trail which appears to lead in the right direction until noon, then turn back if not sure of his location.
E. Stay in the fork in the trail and wait for someone to come by.

In general these items describe a perplexing situation and force the subject to choose the best of several ways of getting out of it. A test of 15 such items had a validity coefficient of .36 in predicting graduation from primary pilot training.

In line with their overall program of test development by factor analysis, AAF psychologists endeavored to find something unique about judgment tests, i.e., some ability brought out by judgment-type items that is absent from other tests. Analysis of these items disclosed that many are correlated with the mechanical-experience factor and many with the arithmetic-reasoning factor. When such items were eliminated and new ones written and analyzed, it turned out that some of the most heavily loaded with the judgment factor were of the work-planning type. These newly constructed items had a loading of .45 on the judgment factor, as compared with a loading of .30 for the items constructed earlier.

Another hypothesis about the nature of the judgment factor led to the development of practical estimation tests. It was noted that numerous items in the judgment tests require the subject to make estimates of the weights of objects, of the time necessary to carry out tasks, and the like. A sample item follows.

Which of the following could be done in the shortest time?

A. Walking 5 miles on snowshoes.
B. Riding a horse 18 miles.
C. Swimming 2 miles.
D. Rowing a boat across a lake 2½ miles wide.
E. Walking 6 miles on flat terrain.

This sample item calls for a relatively simple judgment, as compared to items of the work-planning type. Nevertheless distance, time, and estimates of typical human performance are involved. Items calling for absolute estimates of lengths and sizes were tried out and rejected because of low communality with other judgment tests.

When these practical estimation tests were included in a battery of tests for analysis, they showed some validity for the prediction of success in pilot training and were significantly loaded on the judgment factor. They are not, however, a pure measure of the judgment factor. In addition to their loadings with the judgment factor, they also have significant loadings with the planning and mechanical-experience factors. Moreover, they were found to be less heavily weighted with the judgment factor than practical-judgment items of the work-planning type.

One of the aims of these factor analyses was to raise the validity of tests for the selection of pilots. A moderate degree of success was achieved in this respect, for the test of common-sense judgment had a validity of .17 even when mechanical experience and arithmetic reasoning, for which other good tests were available, had been largely eliminated. This validity, then, is largely unique to judgment tests and will significantly raise the validity coefficients obtained by the use of other tests.

The other aim, more important to the psychology of judgment, was to analyze the nonerror variance of each test into independent factors and from information obtained in this way to find or construct a pure test of good judgment. In this respect the program was unsuccessful since no tests were found that were pure tests of judgment, unrelated to other factors. Although the elimination of mechanical experience and arithmetic reasoning partially purified the test of common-sense judgment, the work-planning and general reasoning factors could not be eliminated.

USC Study of Evaluative Abilities. Some of these leads were followed up by Guilford and his collaborators (Guilford, Hertzka, Berger, and Christensen, *1952;* Guilford, Hertzka, and Christensen, *1953*) in a large-scale study of evaluative abilities. This name was preferred as broader than the usual meaning of judgment to psychologists.

A wide variety of tests, pertinent to several hypotheses about evaluative abilities, were given to about 400 air cadets and student officers. The AAF test of practical judgment or work planning came out on the

general reasoning factor. Several tests intended to test conceptual judgment, e.g., evaluation of syllogisms and inferences, defined a factor which could be called logical evaluation or logical reasoning. No general judgment factor was identified.

From these investigations, and Corter's (*1952*) as well, we must conclude that there is at present no evidence for speaking of a general judgment factor separate from general reasoning.

GOOD JUDGMENT IN CIVILIAN EMPLOYMENT

The nonmilitary use of judgment tests has taken several forms. A Test of Practical Judgment devised by Cardall (*1942*) has items such as the following:

Why are food products advertised in women's magazines?
a. To aid women in planning meals.
b. To make people entertain more.
c. To create a demand for special brands.
d. To make the housewife's work easier.

In various occupational groups scores on this test have low (Cardall, *1942*), medium (Carrington, *1949*), and high (Sartain, *1945*) correlations with general intelligence. In one sample of 46 inspectors in an aircraft factory (Sartain, *1945*), the correlation between this test and criterion ratings was .61, and it contributed somewhat more to a multiple regression equation than a standard test of general intelligence.

A Supervisory Judgment Test has been developed by the United States Civil Service Commission (Mandell, *1947*) for the selection and promotion of foremen in shipyards, air stations, and other field installations of the Navy Department. It has 50 items asking for judgment on problems of employee-supervisor relations and supervisory techniques. When this test was given to 172 foremen in several different installations, the scores were rather closely related to scores from tests of mechanical principles, reading comprehension, and rules and regulations. Its validity coefficients were somewhat higher than those obtained from conventional selection tests (Mandell, *1950*), but there is no evidence that this judgment test contributes anything unique to the test battery.

An interesting Administrative Judgment Test, developed by the Civil Service Commission, has been described by Mandell (*1950a*).

A sample item follows:

Which one of the following administrative situations or problems will most probably occur when direct relations are permitted between a staff specialist employed by the national office of an organization and the operating officials employed in the field offices?

A. decrease in the feeling of responsibility of national office specialists for the operations of state programs in their specialties
B. inadequate technical supervision of field office operations
C. inadequate knowledge in the national office of the competence and qualifications of field office personnel
D. difficulty in keeping the relations on an advisory basis
E. subordination of professional considerations to general administrative responsibilities

Scores on tests of such items have been obtained from several groups of civil service personnel and compared with ratings by colleagues and supervisors. These correlations are around .50, indicating rather high validity. Correlations of tests of general verbal ability with the criterion are substantially lower than these.

We must admit that good judgment is complex; that the construction of a pure test of judgment, independent of other tests, is difficult, perhaps impossible. The important question now, taking these tests as they are, is: What determines success on these tests? The results summarized above furnish at least a partial answer. Those who make better judgments than others often do so because they have more information about the subject matter to begin with. This is particularly true when the judgments require mechanical information or information about distances, sizes, or human performance. Other tests of the same type could be constructed which require information about dates, prices, people's attitudes, habits of enemy commanders, and what not. Or the raw material from which judgments are fabricated could come directly from perception, in which case perceptual abilities play a large part in determining individual differences in the judgments.

For success on some judgment tests more than mastery of the raw material is necessary. The data must be analyzed, weighed, synthesized, or otherwise manipulated. The ability to do this has something in common with the abilities tested by an ordinary intelligence test and much in common with reading and other verbal tests.

In addition to the information and verbal content some judgment tests depend upon the reasoning factor or factors discussed in the preceding section. For practical selection purposes these various abilities

may be combined in the most convenient way, but it is hard to imagine any situation in which the separation of judgment tests from reasoning tests would be worth while.

CREATIVE THINKING

The important questions about individual differences in creative thinking are the same as the questions about reasoning or any other important abilities: identification of the ability or abilities, relations with other abilities and personal characteristics, and relations between this ability and real-life accomplishments.

Unlike reasoning and judgment, which superficially can be identified as abilities, creative thinking as an ability is not readily distinguishable. It is much easier to record creative accomplishment than to identify the abilities responsible for this accomplishment. The study of individual differences in creative thinking begins, therefore, with the criterion rather than with ability measures.

CRITERIA OF CREATIVE ACCOMPLISHMENT

Much has been written about the essence of creativity. It is a favorite topic of lyricists and cynics. For the mundane purposes of this book the starting place is the fact that creative accomplishment by definition must be judged valuable by someone, somewhere, somewhen. There are only a few ways of any significance for empirical research by which such evaluations can be made.

EMINENCE. One criterion of a man's creative ability, perhaps the definitive one, is the winning of a place on the pages of history. The names of Bach, Descartes, Freud, Goethe, and Shakespeare are recorded in the history books, not for excellent routine work, but for imaginative contributions of a high order of originality. It is possible to identify historical geniuses in this way and to compare them with ordinary citizens (see Anastasi and Foley, *1949,* Chapter 17). The difficulty with this procedure, of course, is the paucity of information of psychological significance.

A more modest criterion, applicable to living men and women, is eminence in the eyes of their colleagues. Artists can name the twenty most eminent living artists, for example, and a name mentioned frequently must be an outstanding artist. The scientists listed in *American Men of Science* are eminent by this standard, and usually so because of their creative accomplishments.

Also applicable to living men and women are records of productivity, such as counts of publications and patents. In organizations devoted primarily to research, salary level is another criterion that has been used.

Critical Incidents. The levels of creative ability mentioned above are quite high. In studying other abilities we do not limit our investigation to highly superior thinkers. We have learned from the feebleminded and the brain-injured. There are many occupations in modern society that call for a moderate degree of creative imagination on moderately frequent occasions. Industry employs many engineers assigned to research and development. Commercial art includes a wide variety of occasions for originality. Trainees, including graduate students and research assistants, in many areas of study have opportunities to show originality or to fail to do so. When the milkmaid lured the obstinate calf into the barn by wetting her finger with milk, the philosopher, who had already tried more direct but less successful techniques, noted the incident in his journal as an out-of-the-ordinary contribution to domestic progress.

If critical incidents are to be used as criteria, such occasions must arise rather often, so that originality, or the lack of it, in the employee's work will come to the attention of superiors and colleagues and can be assessed with some reliability. Study of creative accomplishment in these practical situations has many difficulties—which are treated in the textbooks on personnel psychology—but has the advantage of being close to the criterion of real life. In contrast to historical geniuses and eminent men, these men and women who are moderately above average in creative accomplishment are available for comparison with those who are below average.

One careful study of critical incidents involving research personnel has been carried out by the American Institute for Research (Flanagan, *1949*). The personnel investigated were research workers in the Office of Naval Research, the Bureau of Ships, the Bureau of Ordnance, and the Bureau of Aeronautics. All were engaged in basic and applied research in the natural sciences and engineering.

Reports of critical incidents were made on the spot by more experienced research workers. These reports depended on only two types of judgment: first, that the specific activity was of real importance to the job; and second, that it was done so well or so poorly as to warrant special notice. More than 2500 critical incidents were reported, involv-

ing 500 research workers. From these reports the following list of critical requirements was compiled. This list may be taken as a rather complete and systematic arrangement of the activities or criteria that differentiate outstanding scientific investigators from the ordinary.

CRITICAL REQUIREMENTS FOR RESEARCH WORKERS

1. *Formulating Problems and Hypotheses.* Including identifying and exploring problems, defining the problem, and setting up hypotheses.
2. *Planning and Designing the Investigation.* Including collecting background information, setting up assumptions, identifying and controlling important variables, developing systematic and inclusive plans, developing plans for the use of equipment, materials, or techniques, anticipating difficulties, determining the number of observations.
3. *Conducting the Investigation.* Including developing methods, materials, or equipment, applying methods and techniques, modifying planned procedures, applying theory, attending to and checking details, analyzing the data.
4. *Interpreting Research Results.* Including evaluating findings and pointing out implications of data.
5. *Preparing Reports.* Including describing and illustrating work, substantiating procedures and findings, organizing the report, using appropriate style in presenting report.
6. *Administering Research Projects.* Including selecting and training personnel, dealing with subordinates, planning and coördinating the work of groups, making administrative decisions, working with other groups.
7. *Accepting Organizational Responsibility.* Including performing own work, assisting in the work of others, subordinating personal interests, accepting regulations and supervision.
8. *Accepting Personal Responsibility.* Including adapting to associates, adapting to job demands, meeting personal commitments, being fair and ethical, showing interest in work.

Some of these critical requirements appear different from others. Designing the investigation looks different from preparing reports and accepting personal responsibility. One of the important lessons that the psychology of thought can learn from personnel psychology is that different criteria of success are not highly correlated. The correlation between different criteria or different aspects of one criterion is an empirical question, not to be answered a priori. If we study the records of the men included in *American Men of Science* and ask why each was included, we would have to list all of these eight criteria. We could not say at this time just which we would weight most heavily.

We do know, however, that when the critical incidents technique was used in several Navy scientific and engineering laboratories, supervisors emphasized the conduct of the investigation most frequently, both positively and negatively. The formulation of problems and hypotheses was mentioned least frequently, both positively and negatively (Weislogel, *1950b*). The emphasis might be different, of course, in laboratories of pure science.

Work Samples. A third criterion of creative ability is a laboratory procedure, a work sample. We can ask people to write a story, paint a picture, put sticks and blocks together, and so on. Imagination can be demonstrated in almost any assignment in the laboratory, as in real life, but it can be demonstrated most clearly when the instructions call for originality and the problem is so defined that a wide variety of solutions is produced. The advantage of this procedure is that the experimenter has some control over the conditions of the investigation. The disadvantage is that the whole enterprise must be concluded in a short time.

This disadvantage must not be lightly dismissed. Biographies of original thinkers suggest that they carry their assignments around with them for months, even years. Persistence of application to a problem is probably not tested in a two-hour laboratory assignment. On the other hand, standard tests given in an hour or so do predict real-life achievement. And some real-life jobs demand creative output in a few hours. A short-time creative assignment carried out in the laboratory under standard conditions probably is a good criterion of creative ability, but we do not know just how good it is.

The work-sample criterion requires some kind of rating of the product. This is the stumbling block that has discouraged research in this area. The other criteria depend upon ratings also, but they are the ratings of history or of expert colleagues. Laboratory investigation requires that the experimenter be able to rate the merit of the thinker's creative output.

It is possible by careful work to get criterion ratings of work samples that satisfy the requirements of scientific investigation. Andrews (*1930*) stimulated imagination in preschool children by showing them pictures and designs, then found it possible to get reliable ratings for quality of imagination on a scale running from *no imaginative play* to *personification* and *construction*. Unpublished research by the writer, with the aid of Laird Goodman and Rosalind Johnson, demon-

strated that students' stories written during a class period could be rated for originality or literary imagination with a reliability indicated by an interjudge correlation of .76. Such reliability is possible because of the great range of originality in a college class of 50 students.

CORRELATES OF CREATIVE ACCOMPLISHMENT

Admitting that the criteria of creative accomplishment are not very good, several investigators have correlated one or another criterion with other variables of promise.

GENERAL INTELLIGENCE. Many men of ordinary intellect are fond of telling us that geniuses are really pretty dull, or at least that they were dull in childhood. Darwin could not make the grade in school. Einstein cannot add very well. The facts, those that are available, point to the opposite conclusion. Cox (*1926*) was able to gather satisfactory information on the childhood attainments of 300 historical geniuses, which demonstrates that they were far above average in general intelligence. The data consist of records of the ages at which these eminent men read books, wrote letters, and left evidence of other intellectual achievements. By comparing these ages with the ages at which average children carry out equivalent achievements, it was possible to estimate mental age and IQ.

By this historiometric procedure, average IQ for this group of 300 geniuses was estimated at about 160. Allowing for all kinds of error, they must have been far above average in general intelligence. Darwin's childhood IQ was about 150, and several were above 180, such as Goethe, Pascal, and Macaulay.

Even within this eminent group differences were noted. Those who achieved fame as philosophers, poets, dramatists, novelists, and statesmen were highest in estimated IQ, while musicians, artists, and military leaders were lowest. This is understandable in the light of the data presented earlier in this chapter (see page 410), which show that general intelligence is heavily loaded with verbal ability and abstract reasoning.

The relation between intelligence and creative accomplishment has been studied also by beginning with high intelligence. Terman's (Terman and Oden, *1947*) group of children with high IQ's have now reached the age at which some creative accomplishment might be expected. And indeed the group has written many books, many poems, and has been granted many patents. No outstanding geniuses have

been identified, but much creative accomplishment above ordinary levels has been produced.

Several investigators (see Johnson, *1944*) have correlated some measure of imaginative output with scores on a test of general intelligence. The resulting correlations have been positive; they vary in magnitude because of the different samples used, different kinds of imaginative assignments, and different reliabilities of the criteria of imaginative accomplishment.

We must conclude that general intelligence is related to creative accomplishment. Information is lacking to permit a more precise statement of the relationship. Perhaps there is more to be gained by studying group abilities or factors rather than general intelligence. For example, the American Institute of Research has followed up the study of critical requirements for research personnel, described on page 429, by attempting to devise selection tests for each of the eight groups of critical requirements (Weislogel, *1950a*). Tests of formulating problems and evaluating hypotheses have displayed significant correlations with research accomplishments of various groups of physical scientists in government employment (Mandell, *1950b*).

Fluency. Ever since Hargreaves' (*1927*) early study of the faculty of imagination, suggestions have been made from time to time that fluency might be related to originality and creative imagination, the productive noncritical phase of a creative enterprise. Hargreaves scored his fluency tests for originality by counting the number of different ideas, as well as for simple productivity as measured by sheer number of ideas. As Guilford (*1950*) stated the hypothesis recently, "the person who is capable of producing a large number of ideas per unit of time, other things being equal, has a greater chance of having significant ideas."

If there is such a relationship, it is not necessarily a simple relation between ability and accomplishment. If those people who produce the most words or ideas in ten minutes do actually make greater creative contributions, it may be in part because of more interest in the subject matter, more willingness to lose themselves in the task, or more spontaneity of activity. Fluency of imagination is reported to be a good test of the surgent, superficially cheerful personality (Cattell, *1946*). If the picture of the creative man as a cheerful one clashes with the stereotype of the maladjusted genius, it agrees very well with the stereotype of the genius who is happy in his work.

Unfortunately there have been few attempts to verify this attractive hypothesis, probably because of the general belief that a good criterion of creative imagination is hard to obtain. When Primary Mental Abilities tests are correlated with educational achievement, the word fluency test appears to have little predictive value—though this negative finding should not be taken seriously because there is little opportunity for originality on most tests of educational achievement.

A factor analysis of verbal abilities by Carroll (*1941*) is worth consideration as one approach to this topic. Among the 36 tests used were several fluency tests and the writing of a theme on the international situation. The themes written by 119 college students were rated by a number of competent judges in respect to "the excellence of the theme apart from the amount of information exhibited by the subject and the merits of his opinions." After rotation this variable was loaded about .40 on a factor that appeared to be the verbal relations factor—or, as Carroll called it, stock of conventional linguistic responses—and about the same on another factor that was not well identified. There is no good evidence in this factor analysis nor in the table of correlations for the importance of fluency in theme writing.

A report by Dougan, Schiff, and Welch (*1949*) on the use of fluency tests with department store personnel employed in decorating windows offers more encouragement. These four tests, which are called reorganization tests, present unorganized material and ask for more or less meaningful recombinations. The first asks for the construction of short sentences out of ten given words. The second asks for the construction of letters out of two or three given lines. The third asks for the writing of a story from twenty given words. And the fourth requires constructing pieces of furniture from ten blocks. The authors report a contingency coefficient of .60 between scores on this test for 33 employees and supervisors' ratings of the originality of their work. The validity of this test comes in part from its complexity and meaningfulness. It is more closely related to such factors as ideational fluency and verbal versatility (see page 205) than to simple word fluency.

KNOWLEDGE AND EXPERIENCE. Doing something new obviously depends on knowing what is old. Writers are well read. Creative artists are familiar with the successes and failures of older artists and with the standard tricks of their trade. Any scientist's accomplishments are necessarily based on mastery of principles and techniques learned from

other scientists. We may admit that previous learning is not sufficient by itself, but no one will deny that it is necessary.

Quantitative assay of the weight of previous learning relative to other factors is not possible at present. It is worthy of note, however, that in Carroll's (*1941*) factor analysis, mentioned above, originality in writing themes was more closely related to stock of conventional linguistic responses than to any other known factor. Of course, if we restrict our observations to a sample of thinkers who are all well grounded in knowledge from the past, e.g., those who have obtained the Ph. D. in any field of study, the importance of knowledge may be smaller than other factors yet to be identified.

Sensitivity to Social Trends. Highly original work on a problem that no one else is concerned about will never land anyone on the pages of history. Even the most dedicated spirits expect that posterity or the angels will appreciate their contributions. It follows that the thinker who recognizes a social need or trend has a better chance of being judged creative than one who does not.

This ability is clearly necessary for creative accomplishment in applied arts and sciences. Applications are closely related to the demands of society; and when a demand is insistent, discoveries, inventions, and technological advances will be made by different thinkers in different places at the same time (Ogburn, *1922*).

In pure science and mathematics the thinker is not directly influenced by social needs. To be judged creative a thinker of this type must have the ability to attract disciples and to secure favorable judgments from those who support his activities and those who write the histories of his field. In short, he must be an intellectual leader. If the psychoanalysis of Freud or the behaviorism of Hull could be proved completely wrong or socially inconsequential, these men would still be considered eminent creative thinkers because each inspired an intellectual movement of significant dimensions.

In religion, literature, and other arts intellectual leadership is even more important. Some scientific movements can later be evaluated in relation to scientific progress, but in art and religion there is no clear criterion of progress. In these fields intellectual leadership consists in starting a movement that satisfies some social or intellectual need or in getting in front of one that has already started. The thinker who recognizes when an intellectual movement has expired, when the consumers

of his product are ready for a new trend, is in a good position to secure a favorable evaluation of his creative accomplishments.

The creative thinker does not follow the crowd. He must have a certain detachment from the crowd, a sensitivity to the still, small light within. Hence the relationship between sensitivity to social trends and creative accomplishment is not a simple one. The next section will help to clarify this relationship.

PERSONAL CHARACTERISTICS. Through the years many experts have studied the personalities and backgrounds of creative geniuses. Biographers, literary psychologists, and particularly psychoanalysts have interpreted the accomplishments of men of genius as due to industriousness, emotional sensitivity, compensation, neuroticism, rebellion against parents and society, sublimations of various erotic drives, and the like. This department of literature has never suffered from lack of intriguing speculation. Empirical testing of these speculations has been very difficult.

Lacking test data on geniuses and eminent men, several investigators have gleaned as much information as possible from biographical data. Most of the dependable information of this kind has been summarized by Anastasi and Foley (*1949,* Chapter 17). It is clear that eminent men and women come largely from favorable family backgrounds. Their parents are preponderantly well educated and well to do. More grow up in the city than in the country. They are likely to have eminent relatives.

The question of the relation between psychopathology and genius has aroused much controversy. A careful study by Lange-Eichbaum (*1932*) found a large percentage of insanity among historical geniuses. The data are fragmentary, to be sure. The contrary conclusion was reached by Cox (*1926*), following a somewhat more adequate methodology. She was able to get rather good data on childhood behavior of 100 cases and had the records rated for personality and character by two judges. This procedure indicated that children who later became eminent were characterized by persistence, strength of character, emotional balance, and sociality. Likewise, when Shannon (*1947*) searched the biographies of 250 eminent men, mostly scientists, for traits attributed to them, the ones mentioned most frequently were enthusiasm, intelligence, creativeness, *expertise,* and aggressiveness. Last in the list of 46 traits was nonsociability.

Some of the confusion may be cleared up when different kinds of

eminent men are compared. With this end in view White (*1930*) re-examined the data collected by Ellis (1030 cases) and by Cox (300 cases). The two sets of data do not agree altogether, but some trends stand out. In White's words, "it seems that the aesthetic types tend to be more psychopathic than the scientific and practical types. The poets and novelists, in particular, are near the top pretty consistently, while soldiers and statesmen are consistently near the bottom."

Similar results were obtained by Raskin (*1936*) from study of the records of 120 men of science and 123 men of letters. Her summary states, in part:

> Emotionality, despondency and egotism were more frequently reported for literary men than for scientists; the data also indicate for men of letters a definitely more marked tendency toward instability, a higher incidence of ill health in childhood and maturity, a higher incidence of tuberculosis and a slightly shorter life-span. For scientists, the data indicate greater sociality, cheerfulness, and modesty, a low incidence of instability and ill health in childhood and maturity and a slightly longer life-span.

In general we can dismiss the allegation of above-average psychopathology for eminent scientists, but not for artists and literary men. If the aesthetic geniuses are proved to be more unstable, the above discussion of sensitivity to social trends suggests three hypotheses. Accomplishments in arts and letters (1) must be more dramatic, must make a bigger break with social conventions, than scientific accomplishments in order to be judged creative; and (2) are more personal, expressing the creator's own motivations. (3) Because of 1 and 2, society's treatment of the aesthetic genius is more frustrating, and his psychopathology may be a reaction to this frustration. Verification of these hypotheses, or any others, waits for better records than are now available and for better understanding of psychopathic symptomatology.

Putting aside the question of frank psychopathology, it is clear from these studies of biographical records that eminent thinkers are characterized by persistence, desire to excel, and favorable background opportunities. The importance of motivation comes out again in Roe's careful clinical studies of eminent living painters (*1946*), biologists (*1951b*), physical scientists (*1951a*), and psychologists and anthropologists (*1953*). She did not discover any personality or intellectual traits

that clearly distinguished these eminent thinkers from other people, but she noted that they were characterized by unusual energy output and perseverance. Some of the reasons for this energy output could be inferred from study of childhood experiences. There is a suggestion in Roe's material, though the samples are small, that scientists are more conventional in personal habits than artists.

The urge for self-expression, or something like it, comes out also in the attitudes of creative scientists. The more productive scientists of the Armour Research Foundation expressed themselves as more favorable toward self-determination of deadlines, and less favorable toward equalitarian practices in research groups, than their less productive colleagues (Van Zelst and Kerr, *1951*). These personal data are more than psychological curiosa, for a biographical information blank has been found promising in the selection of scientists for government employment (Mandell and Adams, *1948*). For predicting the productivity of social scientists, rapidity of educational progress and age at time of first publication have some validity (Meltzer, *1949*).

ANALYSIS OF CREATIVE ABILITIES

A different approach to this topic has been followed by Guilford and his associates (Guilford, Wilson, Christensen, and Lewis, *1951;* Guilford, Wilson, and Christensen, *1952*). Putting aside for a while the question of validity or relation with known creative accomplishment, they began with hypothetical components of creative thinking, such as sensitivity to problems, fluency, flexibility, and originality. They devised several tests of each of these hypothetical abilities and gave the battery of 44 tests to about 400 air cadets and student officers. The factor analysis disclosed five intriguing new factors:

1. Originality, a bipolar factor with ability to produce uncommon, clever, or remote responses as the positive side;
2. Redefinition, ability to shift the function of objects and use them in a new way;
3. Adaptive flexibility, ability to change set to meet new requirements imposed by changing problems—a bipolar factor;
4. Spontaneous flexibility, ability to change set in a situation in which the direction of set change is unrestricted;
5. Sensitivity to problems, ability to recognize practical problems.

Three fluency factors, word, associational, and ideational, turned up as expected (see page 205).

The originality factor is particularly interesting. Most of the tests of this factor were of the completion or open-end type, the responses being rated by the investigators as to uncommonness, remoteness of association, and cleverness (Wilson, Guilford, and Christensen, *1953*). Plot titles, for example, when rated for cleverness, were the best test of the originality factor, but a simple count of plot titles was the best test of ideational fluency. Incidentally these two ways of scoring plot titles were uncorrelated.

The next best test of originality was a word association test, given quickly, with the responses scored for uncommonness.

The third best test of originality began with questions, such as "What would happen if all national and local laws were suddenly abolished?" The responses scored for remoteness or indirectness came out on the originality factor. A simple count of the responses came out on the ideational fluency factor.

The flexibility factors are not so well defined as the originality factor. Flexibility may not have the same meaning as in Chapter 7.

This ingenious study may mark the beginning of a new trend in the investigation of creative thinking. Fortunately, reference tests were included in the analysis, and these new factors appear to be independent of verbal comprehension, perceptual speed, and general reasoning—in this highly selected sample at least. When these new tests are further refined and the factors more clearly identified, the critical question can be raised: whether they predict actual creative accomplishment.

IMPLICATIONS FOR THE ANALYSIS OF THINKING

The identification of abilities designating the materials of thought, such as the verbal, numerical, and spatial factors, has been fairly successful. The identification of abilities designating thought processes, a more important task for present purposes, has not been so clear-cut. Looking backward through this chapter, it appears that the usual multiple-choice tests, so convenient for large-scale correlational analysis, are easily separated into verbal, numerical, and spatial materials. The multiple-choice item may not be so well adapted to the separation of thought processes, and results to date suggest that more successful differentiation is achieved when the form of the item is changed. The productive process, for example, is tested by fluency tests, quite different from multiple-choice tests.

The apparent success of the USC study of creative thinking in identifying factors that seem to represent processes rather than materials probably depends in large part on the use of items of radically different forms. Other hints for tests of thought processes may be found in experiments summarized earlier, e.g., records of time spent in preparation (see Chapter 6), and in evaluation and revision. The studies by Welch and Long (see pages 231–240) on higher levels of abstraction suggest a way to secure an empirical identification of abstract thinking. The ability to form second-order concepts could perhaps be compared with the ability to form first-order concepts.

Preceding chapters have shown how the difficulty of a problem is influenced by such conditions as emotional involvement, sequence of problems, familiarity of the material, and the context. It should be possible to determine whether different abilities are required when similar test items are given under different conditions.

The problem of timing the tests always raises questions of a balance between convenience and validity. When Myers (*1952*) used speeded and unspeeded reasoning tests, different reasoning factors were identified, and the unspeeded scores were the more valid predictors of grades. Since many of the factors identified by factor analysis come from short speeded tests, this finding has important implications for the naming of the factors. It is true that speeded tests and unspeeded tests are usually correlated, but the correlation is never perfect and the difference in factorial content may be enough to confuse the definition of the factors.

The problem of reliability also enters at this point. Tests of verbal ability, reasoning, and the like, using the conventional type of item, yield scores of satisfactory reliability. Scores for individual differences in some aspect of problem solving, obtained as a by-product of a problem-solving experiment, are often of questionable reliability, e.g., flexibility in shifting set. But if the sophisticated methodology of modern test construction is applied to this difficulty, satisfactory reliability will probably be obtained. In general, what is needed at the moment is a combination of the methodologies of differential psychology and experimental psychology.

REFERENCES

Adkins, D. C., and Lyerly, S. B., 1952. *Factor analysis of reasoning tests,* Chapel Hill, University of North Carolina Press.

Anastasi, A., and Foley, J. P., 1949. *Differential psychology,* New York, The Macmillan Co.

Andrews, E. G., 1930. *The development of imagination in the preschool child,* Iowa City, University of Iowa Studies in Character, 3, No. 4.

Barratt, E. S., 1953. An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors, *J. Psychol., 36:*17–25.

Bernyer, G., 1946–1947. Analyse factorielle de tests d'aptitude technique, *L'Annee Psychologique, 47–48,* 198–207.

Billings, M. I., 1934. Problem-solving in different fields of behavior, *Am. J. Psychol., 46:*259–292.

Blakey, R. I., 1941. A factor analysis of a non-verbal reasoning test, *Educ. Psychol. Meas., 2:*187–198.

Botzum, W. A., 1951. A factorial study of the reasoning and closure factors, *Psychometrika, 16:*361–386.

Burt, C., 1909. The development of reasoning in school children, *J. Exper. Pedagogy, 5:*68–77, 121–127.

Cardall, A. J., 1942. *Test of practical judgment for 12th grade level and above,* Chicago, Science Research Associates.

Carrington, D. H., 1949. Note on the Cardall Practical Judgment Test, *J. Appl. Psychol., 33:*29–30.

Carroll, J. B., 1941. A factor analysis of verbal abilities, *Psychometrika, 6:* 279–307.

Cattell, R. B., 1946. *Description and measurement of personality,* Yonkers, World Book Co.

Corter, H. M., 1952. *Factor analysis of some reasoning tests,* Psychol. Monogr., 66, No. 340.

Cox, C. M., 1926. *Genetic studies of genius: II. The early mental traits of three hundred geniuses,* Stanford University, Calif., Stanford University Press.

Davis, F. B., 1944. Fundamental factors of comprehension in reading, *Psychometrika, 9:*185–197.

Davis, F. B., 1946. A brief comment on Thurstone's note on a reanalysis of Davis' reading tests, *Psychometrika, 11:*249–255.

Davis, F. B., 1947. *The A.A.F. qualifying examination,* Army Air Forces Aviation Psychology Program Research Reports, Report No. 6, Washington, D. C., Government Printing Office.

Dougan, C. P., Schiff, E., and Welch, L., 1949. Originality ratings of department store display department personnel, *J. Appl. Psychol., 33:* 31–35.

Dunning, G. M., 1949. The construction and validation of a test to measure certain aspects of scientific thinking in the area of first year college physics, *Science Educ., 33:*221–235.

Feder, D. D., 1938. Comprehension maturity tests—a new technique in mental measurement, *J. Educ. Psychol., 29*:597–606.

Flanagan, J. C., 1949. Critical requirements for research personnel, Pittsburgh, American Institute for Research.

Gans, R., 1940. A study of critical reading comprehension in the intermediate grades, Teachers College Contributions to Education No. 811, New York, Columbia University.

Glaser, E. M., 1941. An experiment in the development of critical thinking, Teachers College Contributions to Education, No. 843, New York, Columbia University.

Goodman, C. H., 1944. Prediction of college success by means of Thurstone's Primary Abilities Tests, *Educ. Psychol. Meas., 4*:125–140.

Guilford, J. P., 1947. Printed classification tests, Army Air Forces Aviation Psychology Research Program Reports, Report No. 5, Washington, D. C., Government Printing Office.

Guilford, J. P., 1950. Creativity, *Am. Psychol., 5*:444–454.

Guilford, J. P., Comrey, A. L., Green, R. F., and Christensen, P. R., 1950. *A factor-analytic study of reasoning abilities: I. Hypotheses and description of tests,* Reports from the Psychological Laboratory, No. 1, Los Angeles, University of Southern California.

Guilford, J. P., Green, R. F., and Christensen, P. R., 1951. *A factor-analytic study of reasoning abilities: II. Administration of tests and analysis of results,* Reports from the Psychological Laboratory, No. 3, Los Angeles, University of Southern California.

Guilford, J. P., Green, R. F., Christensen, P. R., Hertzka, A. F., and Kettner, N. W., 1952. *A factor-analytic study of Navy reasoning tests with the Air Force air crew classification battery,* Reports from the Psychological Laboratory, No. 6, Los Angeles, University of Southern California.

Guilford, J. P., Hertzka, A. F., and Christensen, P. R., 1953. *A factor-analytic study of evaluative abilities: II. Administration of tests and analysis of results,* Reports from the Psychological Laboratory, No. 9, Los Angeles, University of Southern California.

Guilford, J. P., Wilson, R. C., Christensen, P. R., and Lewis, D. J., 1951. *A factor-analytic study of creative thinking: I. Hypotheses and description of tests,* Reports from the Psychological Laboratory, No. 4, Los Angeles, University of Southern California.

Guilford, J. P., Wilson, R. C., and Christensen, P. R., 1952. *A factor-analytic study of creative thinking: II. Administration of tests and analysis of results,* Reports from the Psychological Laboratory, No. 8, Los Angeles, University of Southern California.

Hargreaves, H. L., 1927. The "faculty" of imagination, *Brit. J. Psychol.,* Monogr. Suppl. No. 10, Cambridge, Cambridge University Press.

Harrell, T. W., 1940. A factor analysis of mechanical ability tests, *Psychometrika, 7*:213–223.

Holzinger, K. J., and Harman, H. H., 1938. Comparison of two factor analyses, *Psychometrika, 3*:45–60.

Holzinger, K. J., and Swineford, F., 1939. *A study in factor analysis: The stability of a bi-factor solution,* University of Chicago, Suppl. Educ. Monogr., No. 48.

Jacobs, R., 1948. The reliability and intercorrelation of the scores on the SRA Primary Mental Abilities Test, *Educ. Res. Bull., 49*:49–58.

Johnson, D. M., 1944. A modern account of problem solving, *Psychol. Bull., 41*:201–229.

Johnson, D. M., and Smith, H. C., 1953. *Democratic leadership in the college classroom,* Psychol. Monog., 67, No. 11.

Jones, L. V., 1949. A factor analysis of the Stanford-Binet at four age levels, *Psychometrika, 14*:299–331.

Kelly, E. L., and Fiske, D. W., 1951. *The prediction of performance in clinical psychology,* Ann Arbor, University of Michigan Press.

Lange-Eichbaum, W., 1932. *The problem of genius,* New York, The Macmillan Co.

Langsam, R. S., 1941. A factorial analysis of reading ability, *J. Exper. Educ., 10*:57–63.

Levine, A. S., 1950. Minnesota Psycho-Analogies Test, *J. Appl. Psychol., 34*:300–305.

Mandell, M. M., 1947. The selection of foremen, *Educ. Psychol. Meas., 7*:385–397.

Mandell, M. M., 1950. Validation of Group Oral Performance Test, *Personnel Psychol., 3*:179–185.

Mandell, M. M., 1950a. The Administrative Judgment Test, *J. Appl. Psychol., 34*:145–147.

Mandell, M. M., 1950b. Measuring originality in the physical sciences, *Educ. Psychol. Meas., 10*:380–385.

Mandell, M. M., and Adams, S., 1948. Selection of physical scientists, *Educ. Psychol. Meas., 8*:575–581.

Matin, L., and Adkins, D. C., A second-order factor analysis of reasoning abilities (to appear in *Psychometrika*).

Meili, R., 1946. L'Analyse de l'intelligence, *Arch. Psychol., 31*:1–69.

Meltzer, B. N., 1949. The productivity of social scientists, *Am. J. Sociol., 55*:25–29.

Miller, W. S., 1947. *Manual for the Miller Analogies Test,* New York, Psychological Corporation.

Myers, C. T., 1952. The factorial composition and validity of differently speeded tests, *Psychometrika, 17*:347–352.

Ogburn, W. F., 1922. *Social change,* New York, B. W. Huebsch.

Pemberton, C., 1952. The closure factors related to other cognitive processes, *Psychometrika, 17*:267–288.

Raskin, E., 1936. Comparison of scientific and literary ability; a biographical study of eminent scientists and men of letters of the nineteenth century, *J. Abnorm. & Social Psychol., 31*:20–35.

Rimoldi, H. J., 1948. A study of some factors related to intelligence, *Psychometrika, 13*:27–47.

Roe, A., 1946. Artists and their work, *J. Personality, 15*:1–49.

Roe, A., 1951a. *A psychological study of eminent physical scientists,* Genet. Psychol. Monogr., No. 43, 121–239.

Roe, A., 1951b. *A psychological study of eminent biologists,* Psychol. Monogr., 65, No. 14.

Roe, A., 1953. *A psychological study of eminent psychologists, and anthropologists, and a comparison with biological and physical scientists,* Psychol. Monogr., 67, No. 352.

Sanua, V., 1952. The contribution of intelligence to improvement in vocabulary, facts and reasoning of psychology, Master's thesis, East Lansing, Mich., Michigan State College.

Sartain, A. Q., 1945. The use of certain standardized tests in the selection of inspectors in an aircraft factory, *J. Cons. Psychol., 9*:234–235.

Shannon, J. R., 1947. Traits of research workers, *J. Educ. Res., 40*:513–521.

Smith, E., and Tyler, R. W., 1942. *Appraising and recording student progress,* New York, Harper & Brothers.

Spearman, C. E., 1927. *The abilities of man,* New York, The Macmillan Co.

Terman, L. M., and Oden, M., 1947. *The gifted child grows up,* Stanford University, Calif., Stanford University Press.

Thorndike, E. L., 1917. Reading as reasoning: A study of mistakes in paragraph reading. *J. Educ. Psychol., 8*:323–332.

Thurstone, L. L., 1938. *Primary mental abilities,* Psychometric Monogr., No. 1, Chicago, University of Chicago Press.

Thurstone, L. L., 1946. Note on a reanalysis of Davis' reading tests, *Psychometrika, 11*:185–188.

Thurstone, L. L., and Thurstone, T. G., 1941. *Factorial studies of intelligence,* Psychometric Monogr., No. 2.

Thurstone, L. L., and Thurstone, T. G., 1943. *The Chicago Tests of Primary Mental Abilities,* Chicago, Science Research Associates.

Thurstone, L. L., and Thurstone, T. G., 1949. *SRA Primary Mental Abilities,* Chicago, Science Research Associates.

Van Zelst, R. H., and Kerr, W. A., 1951. Some correlates of technical and scientific productivity, *J. Abnorm. & Social Psychol., 46:*470–475.

Vernon, M. D., 1947. Different types of perceptual ability, *Brit. J. Psychol., 38:*79–89.

Vernon, P. E., 1950. *The structure of human abilities,* New York, John Wiley & Sons.

Watson, G. B., 1925. *The measurement of fair-mindedness,* Teachers College Contributions to Education, No. 176, New York, Columbia University.

Watson, G. B., and Glaser, E. M., 1949. *Watson-Glaser Critical Thinking Appraisal,* Yonkers, N. Y., World Book Company.

Watters, G. V., and Paterson, D., 1953. Miller Analogies Test scores and ratings of Ph. D.'s in psychology, *Am. Psychol., 8:*89–90.

Weislogel, M. H., 1950a. *The development of a test for selecting research personnel,* Pittsburgh, American Institute for Research.

Weislogel, M. H., 1950b. *Procedures for evaluating research personnel with a performance record of critical incidents,* Pittsburgh, American Institute for Research.

White, R. K., 1930. Note on the psychopathology of genius, *J. Social Psychol., 1:*311–315.

Wilson, R. C., Guilford, J. P., and Christensen, P. R., 1953. The measurement of individual differences in originality, *Psychol. Bull., 50:*362–370.

Chapter 12

Intellectual Activities of Social Groups

> The psychology of social institutions is in large part a real psychology of successive individual thought processes.
>
> —Gardner Murphy

IT IS interesting, and scientifically fruitful, to compare a living individual with a lifeless machine; it is equally interesting, and probably more fruitful in the long run, to compare the problem solving of an individual with that of an integrated group of communicating individuals.

Groups are much less integrated than individuals, but the problem-solving activities of groups are similar in many respects to those of individuals, and the methods of investigation are parallel. The investigation of group activities is more difficult, to be sure, because there is more activity to record. But there is a compensating advantage; all communications between members of the group are overt. Troublesome questions about unobservables, such as contents of consciousness, do not arise because any behavior of an individual that can influence another individual can be observed by the investigator.

There are several general questions to be answered: What do groups do when they attempt to solve problems? What are the determinants of these problem-solving attempts? Are groups superior to individuals? Most of the dependable evidence comes from studies of small groups working under laboratory conditions.[1] Later in this chapter we shall

[1] The early experiments have been conveniently summarized by Dashiell (*1935*) and Murphy, Murphy, and Newcomb (*1937*).

try to put these group activities back into their customary social context.

METHODS OF INVESTIGATION

Group activities have not been studied as long as individual activities. The methodology is not so highly developed. In respect to classification and analysis of functions, however, the study of group problem solving is ahead of the study of individual problem solving.

CLASSIFYING VERBAL CONTRIBUTIONS TO GROUP THINKING

Gross observation of social groups working on a common problem discloses one obvious characteristic of social interaction. The individuals

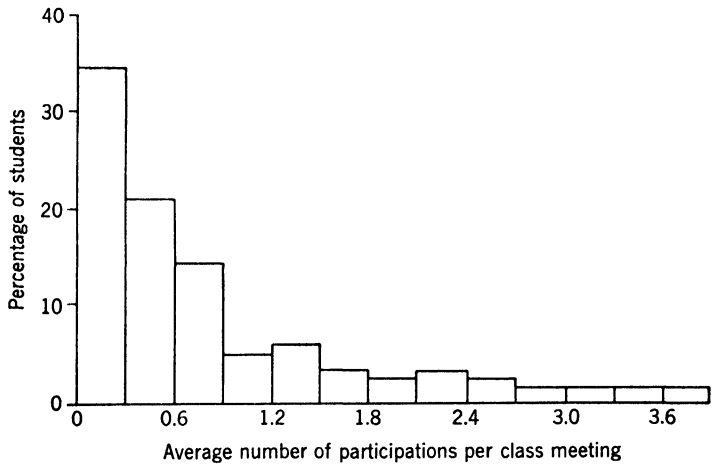


FIGURE 12.1. Distribution of indices of participation in group discussion, four college classes in psychology of about 30 students each. Free discussion was encouraged. Half the students participated less than once in two days. A very few participated more than three times a day. From Smith and Dunbar (*1951*).

of the group do not participate equally. Most members participate seldom; a few talk often. Hence distributions of participation indices are skewed. For an illustration, see Figure 12.1. Beyond this quantitative characteristic it is obvious that different members of the group participate in different ways, at different times. Gross observation fails at this point; refined observation and analysis are required.

Group activities can be analyzed in several ways. Many of the attempts to do this have included all kinds of group activities, without special emphasis on group thinking. Whatever the emphasis, the analysis of the intellectual activities of a group requires inventing and standardizing categories that can be (1) used by different analysts with good agreement and (2) related to the variables or conditions under investigation.

CONFERENCE RESEARCH. The Conference Research project carried on at the University of Michigan (Heyns, *1948*) affords an illustration of the modern approach to the question of reliability or agreement between analysts.[2] Four groups of students from an undergraduate course in the psychology of management discussed an industrial relations problem in hour-long conferences. A psychology professor, not generally known to the participants, acted as discussion leader. Each participant wore a lapel microphone, and an observer kept identifying data in shorthand. Extensive records in typescript constitute the data for analysis.

After preliminary analysis the categories shown in Table 12.1, designating problem-solving functions, were used. The percentages

TABLE 12.1. Problem-Solving Functions and Their Frequency[a]

Functions	Percent Frequency
1. Goal setting	2.9
2. Problem proposals	1.5
3. Information seeking	2.9
4. Information giving	5.3
5. Solution proposals	10.1
6. Development seeking	11.3
7. Development giving	33.3
8. Opposing	17.9
9. Supporting	11.9
10. Summarizing seeking	.4
11. Summarizing giving	2.1
12. Nonproblem-directed	.4

[a] From Michigan Conference Research (Heyns, *1948*).

shown in the table are for the frequency of each function in four one-hour conferences of ten people each, in which 2131 contributions were made.

The first task in coding the typescript into these categories or func-

[2] For a general description of this project, see Marquis, Guetzkow, and Heyns (*1951*).

tions is unitizing (Guetzkow, *1950*). A unit was defined as a subject-predicate unit classifiable into a single functional category. Three reliability checks on the boundaries of the units indicated agreement between analysts, after training, of 76, 90, and 92 percent.

Reliability of coding the groups' contributions into these categories can be computed in several ways. The proper way is a test of the reliability of the scores or percentages in the form in which they are to be used. If the aim is to describe a group's problem-solving activities in respect to the percentages of the contributions of each kind, the reliability of these percentages is the critical figure. In this sense the average discrepancy between the percentages returned by two coders was about 1.5 per cent. If the aim is to describe a group's activities in respect to the rank of the frequencies of each kind of contribution, the reliability of this rank order should be computed. Correlations between such rank orders for two coders ranged from .81 to .98. If the aim is to describe members of the group in respect to percentages of contributions of each kind, the reliability of such percentages should be computed; and these also were high.

In general, this Conference Research project demonstrated that the problem-solving activities of small groups can be recorded and classified into these 12 categories by trained analysts with high reliability. An argument for the validity of the system is the finding that groups that produced solutions of high quality did more solution proposing and development giving than the average group.

Interaction Process Analysis. Another interesting list of categories has been described and used by Bales (*1951*) for the study of social interaction. One group, for example, consisted of two graduate students and four staff members meeting for three hours to discuss the thesis plans of one of the students. The percentages in Table 12.2 are for the pooled data (2616 contributions in all) from three such groups.

In Bales's method the discussion is picked up by microphones and recorded, but, in addition, trained observers listen in and code the discussion continuously as it occurs. The recording is needed only for verification. The unit is the smallest discriminable segment that the observer can classify. Reliability may be computed in several ways. Bales (*1952*) says that "with competent observers and hard training, correlations between observers ranging from .75 to .95 can be obtained."

A special feature of Bales's system is the consolidation of the categories. The list of 12 categories is symmetrical: 6 and 7 go together

as most relevant to problems of orientation; 5 and 8, problems of evaluation; 4 and 9, control; 3 and 10, decision; 2 and 11, tension reduction; 1 and 12, reintegration. It is these larger categories that are related to other variables.

TABLE 12.2. Problem-Solving Functions and Their Frequency[a]

Functions	Percent Frequency
1. Shows solidarity	.6
2. Shows tension release	6.9
3. Agrees	9.7
4. Gives suggestion	4.6
5. Gives opinion	31.9
6. Gives orientation	27.2
7. Asks for orientation	5.5
8. Asks for opinion	4.4
9. Asks for suggestion	1.6
10. Disagrees	4.0
11. Shows tension	2.2
12. Shows antagonism	1.2

[a] From Interaction Process Analysis (Bales, *1951*).

These two methods of analysis may be taken as illustrative of sophisticated attempts to record and classify all or nearly all of a group's intellectual activities. Certainly they have much in common. Both are more refined than the rough trifunctional analysis used in earlier chapters for intellectual activities of individuals. Other investigators have used still different systems of classification. Comparison of the merits of different systems would have to consider reliability, utility for the purpose of the moment, and, above all, learning time. For some comparisons between groups, ratings of a group's insight, productivity, or flexibility may be sufficient. Categorizations and ratings give approximately the same picture of group characteristics (Carter *et al.*, *1951*).

Other investigators have recorded only one or two group activities of particular interest: suggested solutions and rejections of these, words or nonsense syllables recalled, judgments of simple stimuli. These activities are clearly identifiable as such; questions of reliability of observation presumably do not arise. And of course some investigators have simply timed the group's deliberations and recorded the outcome as a success or failure.

GROUP ACTIVITIES AS DEPENDENT VARIABLES

Experimental analysis, as applied to this topic, requires the variation of experimental conditions and determination of their effects on group activities. Most studies of this kind have investigated single items of group behavior, such as judgments of the length of a line or kinds of solution to a problem. In some experiments, however, as in the Conference Research mentioned above, the dependent variables have been the frequencies of all the group processes resulting from the analysis.

CORRELATIONAL ANALYSIS

If a large number of groups can be studied, correlations can be computed between various items of each group's behavior, analogous to correlations between scores for individuals on various tests. Studies of this kind to date have been concerned chiefly with nonintellectual traits (Cattell and Wispe, *1948;* Hemphill, *1949*), but the technique is equally applicable to analysis of the intellectual traits of small groups.

SOME DETERMINANTS OF GROUP PROBLEM SOLVING

What determines the activities of a group attempting to solve a complicated problem? There must be many factors and conditions that influence a group's intellectual processes, only a few of which have been isolated and experimentally examined.

THE COURSE OF GROUP PROBLEM SOLVING: SEQUENCES

Early studies of the thought processes of the individual thinker attempted to describe sequences of functions or stages in progress toward a solution. The assumption was made on logical grounds that some processes, e.g., preparation, orientation, had to be performed before others, e.g., evaluation, could be performed. Hence temporal analysis should yield high frequencies for some activities early in the course of problem solving and high frequencies for other activities later. The outcome of these attempts has not been encouraging, except in the limited respects summarized in Chapter 2. The difficulty is that problem solving has a wheels-within-wheels or reflexive character. Any minor activity, even preparatory activity, may be halted and evaluated. In a complicated problem the final judgment may be preceded by a separate act of preparation. Hence preponderances may be discovered but not clear-cut sequences. Parallel attempts at discovery

of sequences in group problem solving have been undertaken and parallel difficulties have arisen.

The Michigan Conference Research project (Heyns, *1948*) took a cue from John Dewey and tried to find logical sequences of problem-solving functions. For example, in the "felt difficulty" stage there might be relatively high proportions of problem proposals. But when the records were broken down by time segments, e.g., by 15-minute periods, every period was much like every other period. The attempt was also made to uncover optimal sequences, of major or minor magnitude, with only limited success. A minor sequence noted was solution proposal, followed by development seeking or development giving, followed by opposing or supporting. But in all conferences less than 3 percent of the three-contribution sequences were optimal in this sense.

When the sequences were reduced to pairs and tabulated without regard to any optimum, the following pairs of contributions were most frequent:

Development giving, development giving
Development giving, supporting
Supporting, opposing
Opposing, opposing
Development giving, development seeking

From Bales's data (*1951*, page 129) of 2616 contributions of three groups, five similar pairs of most frequent occurrence may be extracted.

Gives opinion, gives opinion
Gives orientation, gives orientation
Agrees, gives opinion
Asks for orientation, gives orientation
Shows tension release, shows tension release

The fact that giving an opinion follows giving an opinion is important but must be interpreted in the light of overall frequencies. Giving an opinion and giving orientation were by far the most frequent activities. Bales, therefore, raised the question of dependent probability: How much more frequently, or less frequently, does one activity follow another than would be expected from the overall frequency of the second activity? For this analysis the activities were grouped into larger categories, and the results permit such a statement for these three groups as: asking for orientation, opinion, or suggestion

(Categories 7, 8, 9) is followed less frequently than expected by showing solidarity, tension release, or agreement (Categories 1, 2, 3) or the reverse of these (Categories 10, 11, 12) and is followed more frequently than expected by giving suggestion, opinion, or orientation (Categories 4, 5, 6).

Another type of analysis of sequences, reported by Bales and Strodtbeck (*1951*), is shown in Figure 12.2. Results obtained from observa-

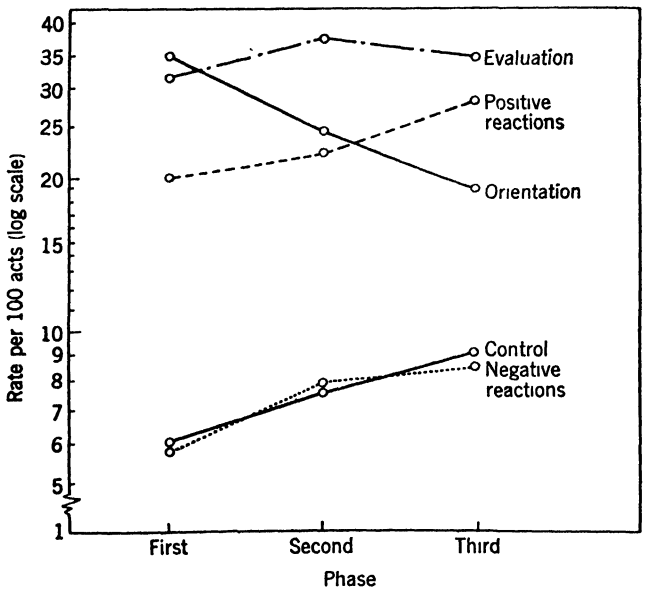


FIGURE 12.2. Relative frequencies of types of interaction in 22 discussions divided into thirds. From Bales and Strodtbeck (*1951*).

tion of 22 discussions of such problems as planning a Christmas party and the first move of a chess game were divided into thirds. From the first third to the last third, orientation activities decreased in frequency; positive, negative, and control activities increased. In general Bales and Strodtbeck suggest that "the interaction process should be considered as a system, with internal tendencies which make each part of the process a condition to other parts."

The evidence for these systematic nonrandom sequences was most significant for seven (out of 22) discussions which met certain speci-

fied criteria for a "full-fledged" problem. For example, if the members of the group have full information about the problem, the problem is "truncated," and there will be little need for giving orientation. The question now becomes one of describing the conditions under which these criteria will be fulfilled and determining whether "full-fledged" group problems are common or rare.

THE STRUCTURE OF COMMUNICATIONS

The groups described above enjoyed complete communication; anyone could talk with anyone else. This may be an optimal condition of communication, but it is probably not the most common. In most social groups the channels of communication are limited by physical conditions, social conventions, rules imposed by a leader or superior group, or in some other way. What happens to group problem-solving activities when the channels of communication are experimentally limited?

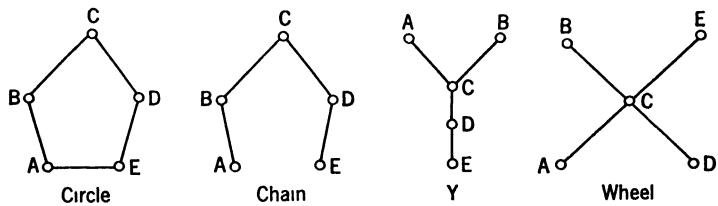


FIGURE 12.3. Communication patterns used by Leavitt (*1951*) in studies of problem solving by groups of five. The wheel was the most efficient, the circle the most enjoyable.

In order to control communication Leavitt (*1951*) separated his subjects by partitions through which there were passageways that the experimenter could open or close. The communication patterns available to groups of five subjects, called circle, chain, Y, and wheel, are shown in Figure 12.3.

The problems that Leavitt put to his groups required exchange of information. Each subject passed notes through an open passageway asking for the information he needed. The groups worked continuously until every member had the facts he wanted and was able to solve the problem.

When group activities were analyzed, it appeared that in the wheel pattern the peripheral men funneled information to the center man, C,

who made decisions and returned answers. In the Y pattern the center man, C, made the decisions, while D transmitted information both ways. In the chain information was channeled from both ends to C, and answers returned both ways. There was no consistent organization in the circle; messages were sent and answers returned in both directions.

All groups improved as they worked successive problems. In general, the order of decreasing efficiency was wheel, Y, chain, circle. As to enjoyment of the task, the order was reversed. These college students

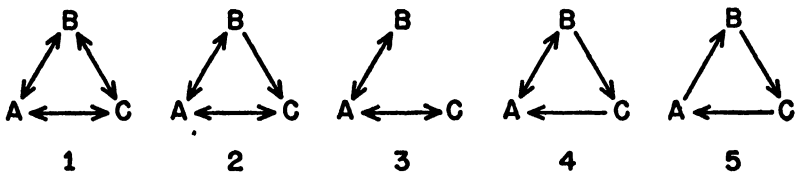


FIGURE 12.4. Communication patterns used by Heise and Miller (*1951*) in studies of problem solving by groups of three. The arrows indicate the direction of communication from talker to listener. Pattern 1, permitting two-way communication for all subjects, was the most efficient for problems solved by unorganized stereotyped communication and was susceptible to noise in the communication system. Pattern 3, characterized by "centrality," was most efficient for problems that require coördination of group thinking. Pattern 5, permitting only one-way communication, was inefficient and very susceptible to noise.

enjoyed the circle most, for they all had a chance to participate and make decisions. They disliked the wheel, except the man who was in the center having all the fun. Apparently the characteristic of communication patterns that Leavitt calls "centrality" increases effectiveness of group function, and simultaneously limits independence and satisfaction of the members.

Heise and Miller (*1951*) controlled the communications of their subjects by locating each in a separate room with a microphone and earphones. Each group consisted of three members. There were five communications patterns, ranging from one extreme in which each member enjoyed two-way communication to the other extreme in which A could talk only to B, B only to C, and C only to A (see Figure 12.4). Additional difficulties were introduced by putting various amounts of random noise into the communication system. The effects

of these arrangements on problem-solving efficiency, as measured by number of words communicated and time required for solution, depend on the nature of the problem.

The first problem called for a comparatively stereotyped and unimaginative exchange of information about words. For this purpose Pattern 5, with only one-way communication, was the least efficient. Pattern 1, with complete two-way communication, was most efficient. Since communication was important, noise reduced efficiency, especially in Pattern 5.

The second problem was less clearly structured and required more initiative and coördination. The group had to reassemble a sentence, the words of which had been distributed to the members of the group. Again one-way communication was less efficient, but Pattern 3, with more of what Leavitt called "centrality," was more efficient than Pattern 1. As in the first problem, noise reduced efficiency, especially in the one-way pattern.

The third task was a kind of anagram game in which each subject was to make as many words as possible out of given words. The score for the group was the total for the group even if all had written the same words. In this case communication is relatively unimportant and, as expected, the communication nets did not differ in efficiency. Noise lowered group scores slightly.

We must conclude from these two ingenious experiments that comparisons between group structures cannot ignore the nature of the problem. Some problems can be solved in an additive way, by relatively unorganized communication. For these problems free two-way communication is helpful. Problems that depend on organized communication are solved better by a group structure which places one member in a central position of leadership to channel communication more systematically. The gain from organization appears in groups as small as three; presumably it would be greater in larger groups. When the information communicated is critical and there is little group redundancy, as in the one-way patterns, any interference with communication will naturally reduce group efficiency.

DYNAMICS OF THE GROUP

The first experiments on intellectual activities of groups were probably too intellectual. They assumed that all members of the group were trying to solve the problem the experimenter asked them to solve.

That may be true in a general way, but each individual brings along his own problems and his own motives when he enters a social group. Preceding chapters have shown how subtle dynamic factors absorbed from the social atmosphere and the experimenter's instructions slant the individual's thinking one way or another. When the individual becomes involved in group activities, he retains these personal dynamics —and additional group dynamics may be built up in the group situation.

Can an observer really detect whether an individual's contribution to group thinking is directed toward the individual's goal or the group's goal? Fouriezos, Hutt, and Guetzkow (*1950*) have shown that ratings of self-oriented needs can be reliably made from observations of conference behavior, and that such ratings agree to some extent with ratings made from clinical data. In decision-making groups studied by the Michigan Conference Research project, groups that were highly self-oriented were somewhat below average on satisfaction with group decisions and on objective measures of group productivity.

In the same project (Marquis, Guetzkow, and Heyns, *1951*) it was found that productivity of a group was related to the urgency of the problem and the power of the group to deal with it. Cohesiveness of the group seemed to be unimportant.

Several attempts have been made to arrange social climates that emphasize one kind of group thinking or another. The tradition limits the effectiveness of these experimental arrangements, but artificially coöperative groups can be formed within a traditionally competitive situation. Deutsch (*1949*) divided college students into small groups, and encouraged coöperation in five of these by promising rewards on a group basis and competition in the other five by promising rewards on an individual basis. According to ratings by outside observers, the coöperative groups displayed more productivity and insight during discussion of all problems. A puzzle problem, which had a clear-cut solution, was solved more rapidly by the coöperative groups. On the other problems, solutions to which were rated by three judges, the coöperative groups produced solutions of higher quality.

Mintz (*1951*) gave his groups a bottleneck problem, ingeniously devised so that coöperative attack was necessary, in order to study the effects of motivation and emotion. Each subject had to pull a cone out of the bottle, but the neck was so small that two cones could not be

withdrawn at once. Groups of 15 to 21 subjects could pull their strings at will or they could plan a coöperative solution.

When the task was presented as simply one of coöperative planning, the groups succeeded quite well, all such groups withdrawing all their cones in a minute or less. In some experiments accomplices made noise and stirred up general excitement, but such disturbances lengthened total times very little. Some other groups were specially motivated by small individual payments and fines. In most of these, serious traffic jams developed at the bottleneck, so half the cones were still in the bottle at the time limit. Hence Mintz attributes the nonadaptive behavior of some excited groups and crowds not to the emotion *per se* but to the secondary effect of emotion when it prevents coöperative planning. Three of the payment-and-fine groups were interrupted in the preliminary planning stage, and all of these later developed serious traffic jams.

An experiment by Wright (*1942*) illuminates the joint effects of social and emotional factors on children's constructiveness as they played in groups of two. The subjects were preschool children: 18 pairs of strong friends, and 21 pairs of weak friends. Constructiveness was rated, with fair reliability, on a four-point scale from 1 (toys used as objects to be shaken or swung without regard to their particular character) to 4 (planned play, such as driving a truckload of cattle to Chicago and selling it). The children played under ordinary conditions and also after being frustrated by deprivation of attractive playthings.

When the children played together, they were more constructive than when each played alone. When they were frustrated, their constructiveness decreased; but strong friends were consistently more constructive, both in free play and after frustration, than weak friends. For both strong and weak friends the biggest effect of frustration was shown in solitary play. These results are consistent with the assumption that social support gives the security needed for constructiveness and for control over frustration.

Several experimenters, beginning with South (*1927*), have noted that all-male and all-female groups do better than mixed groups. There are at least two possible reasons for this difference: more good friends in the like-sex groups, more distractions and instability in the mixed groups. The social activities in the two kinds of group have not been carefully compared.

LEADERSHIP AND GROUP ACTIVITIES

Most natural groups are organized around a leader, either a leader appointed externally or one chosen by the group. This leader may influence both the structure of the group and its dynamics, so his role in group problem solving can be critical.

The Michigan Conference Research had the leader shift his behavior in different groups so that his influence on group processes could be analyzed. In two groups he was to act in a positive manner, facilitating communication among participants and promoting congenial interpersonal relations. In two other groups he was to act in a negative manner, misinterpreting and deliberately misunderstanding, interrupting and promoting irrelevant comments. (Analysis of the observers' records showed that the leader did succeed in following these instructions.) When the leader acted positively, the participants did more supporting and development seeking and less opposing than when he acted negatively. This verifies a common hypothesis about group processes.

Two other hypotheses were checked. One states that the members of a group reflect the leader's behavior. If he increases his activity in one category, e.g., development giving, the group will do likewise. No evidence for this hypothesis was found. An opposing hypothesis states that the group fills in the gaps in the leader's behavior. If he increases some activities, the group reduces its contributions of this kind. This hypothesis also found no support.

Sooner or later any discussion of leadership encounters the question of the influence of the leader who has expert knowledge. Maier (*1950*) considers the case of the leader who has an "elegant solution" for an industrial relations problem and wants a group to accept it. The best way, Maier says, is to guide the group, by good leadership, to discover this solution for itself. The important feature of good leadership in this case is to help the group change "direction" (see page 72) by avoiding frustration and fixation on poor solutions.

When college students, working as individuals, were asked to solve this industrial relations problem, no elegant solutions were produced. When groups of college students were asked to play the roles of the workers in this situation, one being the foreman, and solve the problem by group discussion, a few elegant solutions emerged. When groups were led by an untrained leader who knew the elegant solution, a few more elegant solutions were reported. And when a trained

leader who knew the solution guided the discussions, 12 out of 17 groups produced and adopted the elegant solution. Maier's conclusion is that the leader "can learn to conduct conferences in such a manner as to stimulate thinking and thereby have his ideas rediscovered and accepted."

An important phase of the leader's function that should not be overlooked is the preparation for the group's activities. Ordinarily the leader arranges the meeting, prepares the agenda, and formulates the problem. Thus it sometimes turns out, in real-life groups and in laboratory groups, that the problems posed by the leader (or experimenter) have no personal significance for the members of the group. Casual observation suggests that they often reformulate the topic under discussion so as to give it more personal significance. We cannot say in advance of experimental study whether groups work more effectively on impersonal problems or on problems in which they are personally involved. Like individual thinkers, groups think more persistently about their own problems, but they also get more excited and often more rigid. Group goals may be replaced by individual goals. Should the leader formulate the problem so that it is received as the group's own problem? Or should he formulate it so that the group handles it with cool objectivity?

SOME DETERMINANTS OF GROUP JUDGMENTS

The judgment process taken alone is simpler than the complete process of problem solving, which includes preparation, production, and judgment. Because judgment eventuates in a simple response, one of a small number of categories of response, it is somewhat easier to study this response as a dependent variable and to determine the conditions leading up to it.

THE COURSE OF GROUP JUDGMENT

A small social group that makes important decisions is a jury in a court of law. The actual deliberations of a jury cannot be studied by an outside scientist, but two approximations to this ideal have been carried out. Weld and Roff (*1938*) used the actual records of the Thomas Hoag Case, a celebrated criminal trial in which the defendant was charged with bigamy. The testimony was divided into sections and read to college upperclassmen preparing for the law. After each

section of testimony the subjects indicated their judgments on a scale from 1 (certainty of innocence) to 9 (certainty of guilt).

Figure 12.5 shows how the median judgment of a group of 50 subjects went up when the prosecution presented its evidence and down when the defense replied. The two portions of Figure 12.6 were obtained by presenting the same testimony in other, artificial sequences to new groups of 50. By comparing Figures 12.5 and 12.6 it is clear that the sequence of testimony does change the final judgment signif-

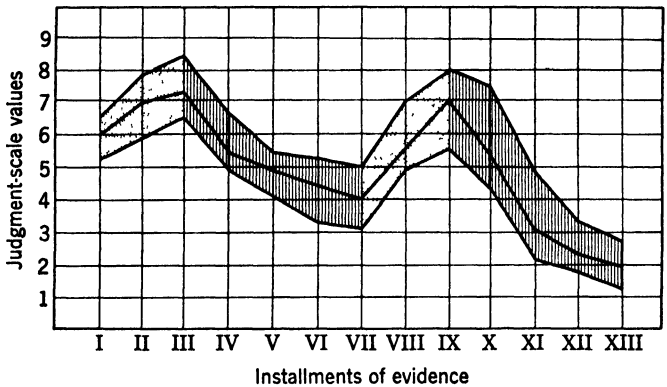


FIGURE 12.5. The relation between group judgments and installments of testimony. The center line represents the median value of all the judgments for each installment of testimony. The width of the figure represents the quartile deviation above and below the median. The stippled portion shows the effect of the evidence for the prosecution; the barred portion, that for the defense. From Weld and Roff (*1938*).

icantly. Analysis of individual differences indicated that some subjects tried to maintain an attitude of doubt, while others were easily swayed one way or the other by new evidence.

Weld and Danzig (*1940*) took advantage of a trial in a moot court conducted in a realistic manner by a professor of law. Three juries of 12 to 17 psychology students recorded their judgments serially while listening to the evidence in the courtroom, then retired to separate jury rooms for discussion and voting. As before, the judgments went up and down with the testimony. The opening and closing statements

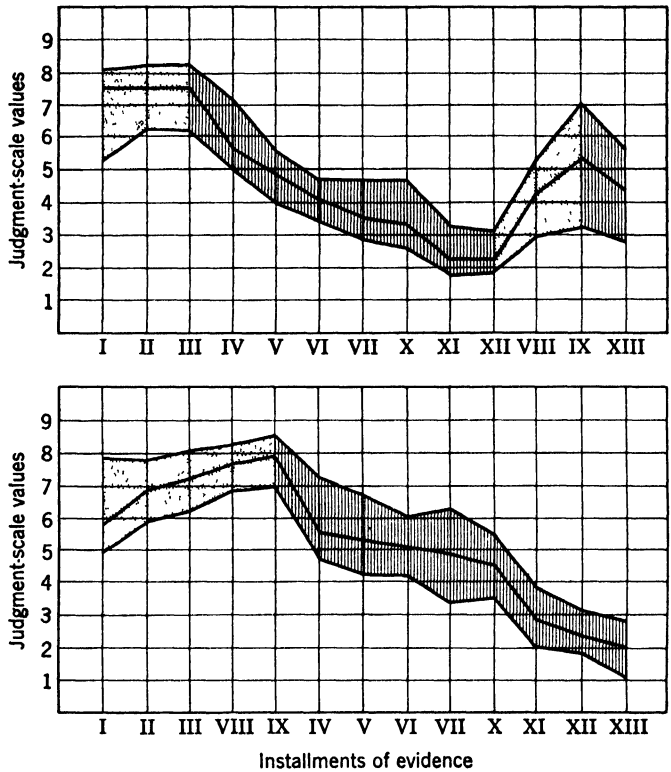


FIGURE 12.6. Effects of variations in sequence of testimony. In these experiments the testimony was the same as in Figure 12.5, but the sequence of the installments was varied. From Weld and Roff (*1938*).

of the attorneys were particularly effective. All jurors but one had their decisions made when they entered the jury room and were not moved by the arguments of their fellow jurors.

TREND TOWARD GROUP HOMOGENEITY

Earlier chapters have shown how social forces influence the individual's judgments. There is another trend that cannot be described in reference to the individual judgments, but must consider the group as a whole. A series of experiments by Jenness (*1932*) show that social factors can produce a decrease in the range of the judgments. He

simply had college students estimate the number of beans in a bottle, then, after various kinds of social interaction, estimate again. In some of the experiments the second group median was more accurate than the first and in some it was less accurate, but in all experiments the trend toward homogeneity was clear-cut.

One group of college students made their first estimates individually, then were separated into committees of three each in such a way that there would be wide diversity of opinion in each committee. That is, a committee might have one member who had guessed high, one who had guessed medium, and one who had guessed low. These committees met in separate rooms and discussed the matter, then all returned to the classroom and made their second estimates. The range of the estimates dropped sharply from the first to the second, as did the standard deviation and the quartile deviation (see Table 12.3).

TABLE 12.3. Effects of Social Interaction, in Groups of Three Degrees of Diversity of Opinion, on Judgments of the Number of Beans in a Bottle (There Were 811) [a]

Composition of groups	N	Group Medians		Quartile Deviations	
		1st	2nd	1st	2nd
1. Maximum diversity	26	799	758	225	106
2. Minimum diversity	24	875	773	216	161
3. Natural groupings	23	750	750	337	125
4. Control: no groups	28	916	925	166	197

[a] Data from Jenness *(1932)*.

Another group went through the same procedure except that the committees were made up, secretly of course, so that there was minimal diversity of opinion in each committee. On the second estimate the measures of group deviation were moderately lower.

In the third experiment the subjects were allowed to group themselves naturally, and under these conditions there was a very strong trend toward homogeneity of judgment, shown by a large reduction in the quartile deviation.

As a control, one class of students also made two estimates, but had

no knowledge of others' estimates. Under these conditions there was a negligible increase in the quartile deviation. Hence the trend toward homogeneity in the group must be due to social interaction of some kind, either the discussion of the assignment or merely the knowledge of others' opinions.

In Sherif's (*1935*) experiments on the autokinetic effect the same tendency toward convergence of judgments was noticed. The final opinion of the group could be called a social norm, and this phenomenon could account in part for the development of common standards in large social groups. Further verification has been added by Bovard (*1951*) for classroom judgments of the lengths of rectangles, and there was some evidence that the convergence was greater in group-centered classes than in leader-centered classes.

This convergence of estimates or trend toward group homogeneity after discussion must be a general phenomenon of group thinking. If the judgments are made in two categories, this phenomenon will be manifested as a trend in the percentage of judgments away from a fifty-fifty split toward a preponderance of one category.

ARE GROUPS SUPERIOR TO INDIVIDUALS?

This question takes several forms. Can a group solve a problem that the individuals of the group cannot solve? Or, if both groups and individuals can solve it, can the group solve it better or faster or with greater efficiency?

We shall see that these questions have different answers for different kinds of problems. We can make use of the trifunctional analysis of problem solving used in earlier chapters. There are no comparisons for problems of preparation, but groups and individuals have been compared on problems that emphasize production and on problems that emphasize judgment.

PROBLEMS OF PRODUCTION

Let us begin with some logical considerations. The anagram problem may be taken as typical of problems that depend principally on production of ideas, rather than preparation or judgment. It requires the production of short words constructed from the letters of a long word. If a small group coöperates on such a problem and the score is simply the number of words produced, the group will of course do

better than the individuals because the group score is the sum of the individual scores. If the score is the number of *different* words produced, the group score will be larger than any individual score whenever there is any variation between individuals in the specific words produced. Consequently, if the problem is one that depends in a straightforward way on the production of certain critical words or ideas, we may infer that two heads will be better than one.

If we are interested in efficiency rather than gross production and we measure efficiency in number of words produced per man-minute, the efficiency of the group is an average of the efficiencies of its members. If we calculate efficiency as number of *different* words produced per man-minute, the group will necessarily be inferior to the average individual whenever there is any overlapping of words produced by the individuals.

These are a priori considerations that have nothing in particular to do with social interaction. If we wish to know what social interaction adds to or subtracts from individual production of ideas, we must begin with these logical considerations as a baseline.

Thus in Watson's (*1928*) experiment with anagram problems coöperating groups produced 75 words in ten minutes, while the members of these groups, working individually in a control experiment, produced only 32 words on the average. But the different words produced by the members working individually total 87. This is what one would predict from the above a priori statements, and it is impossible to say how much social processes encouraged, censored, or slanted the individuals' productive efforts. The importance of the individuals' contributions to Watson's groups is underlined by the correlation of .53 between outputs of different groups and outputs of the best individuals in these groups.

The anagram task, in which individual contributions are simply added, with due regard for duplications, is a special case of group thinking. In other tasks individual contributions may perform any of the nonadditive functions listed at the beginning of this chapter: goal setting, development seeking, summarizing. The individual contribution is embedded in the flow of group activity and becomes difficult to evaluate separately. In such cases logical analysis is not enough; empirical comparisons are necessary.

Watson was able to compare a variety of tasks by having each task performed by individuals and by groups (summarized by Murphy,

Murphy, and Newcomb, *1937*, pages 730–734). The following list ranks the tasks in the order of group superiority of achievement over individual achievement. The rank for group superiority in speed was similar, except that finding opposites drops to seventh place.

Finding words meaning opposite of given words
Solving a cipher
Drawing conclusions from given facts
Sentence completion
Listing steps to solve a problem
Composing limerick
Comparison of numbers
Reading comprehension
Original intelligence problem.

Somewhat similar results were obtained by Husband (*1940*) with college students working alone and in pairs. The pairs were faster in solving a jigsaw puzzle and word puzzles, but about the same as individuals in solving arithmetic problems. Pairs who were friends did better than pairs who were strangers—as in Wright's study of young children. The pairs were not twice as fast as the individuals, so in terms of man-minutes we would have to conclude that the individuals working alone were more efficient.

Shaw (*1932*) used complicated problems involving many steps in sequence, such as transferring missionaries and cannibals across a river under certain limitations. In these serial problems groups of four were clearly superior to, but not more efficient than, individuals working alone in respect to number of correct solutions.

The social interaction of the groups working these complicated problems was not recorded completely, but an observer in each group recorded two kinds of activity: the production of suggested steps or hypotheses, and the acceptance or rejection of these. The records showed that when a suggestion was rejected, it was usually rejected by someone other than the one who made it. This give and take is characteristic of social thinking; if it is oriented toward the problem rather than toward the individual, it may be an important factor in the superiority of group thinking.

Another kind of problem that emphasizes the production of possible solutions is the familiar parlor game called Twenty Questions. Someone thinks of an object which others are to identify by asking questions, usually proceeding from the general to the specific. "Is it

edible?" "Is it a kind of meat?" "Is it lamb stew?" College students working in groups of two were superior to individuals in terms of number of questions needed, number of failures, and elapsed time. Groups of four were superior to groups of two only in number of failures (Taylor and Faust, *1952*). In respect to man-minutes of work the individuals were more efficient. In this type of problem the advantage of group thinking cannot lie merely in the addition of hypotheses to be tested by questions, because the number of questions is limited. It is more likely, as Shaw's results indicated, that an unproductive approach will be recognized by someone in a group than by the solitary thinker.

The importance of the nature of the problem is emphasized again by McCurdy and Lambert (*1952*). They designed an apparatus, like the familiar Yerkes multiple-choice apparatus, which can be operated by one, two, or three subjects. The problem is to turn on a light by manipulating six two-position switches. The single operator has all six switches; each member of a three-man team has two switches. Each of the three members of a team must operate his two switches in the right way at the right time in order to match the plan of the experimenter's manipulations. In this kind of problem each individual must participate, and an incompetent individual can drag down the group performance—though one individual could do all the planning if he could get the others to follow him. The attempts are just guesses at first, but the plan can be discovered.

Two experiments with this apparatus problem, one of 11 individuals compared to 13 three-man groups, and one of 12 individuals compared to 22 three-man groups, showed a slight superiority for the individuals. The groups made more errors at the beginning, probably because there is often one member of the group who does not attend to the instructions.

Further investigations with this apparatus problem analyzed the effects of communication and of leadership. Groups that were restricted to gestural communication did not do as well as groups that used both gestures and words. Groups that had a leader, appointed by the experimenter, were no better or worse than groups operating without a leader.

All these experiments, taken together, emphasize two factors in comparisons between individual and group thinking: the nature of the problem, and the variability of the members of the group. Group

thinking looks good when the problem is one that can be solved by the addition of individual contributions and by group correction of individual errors. Group variability increases the variety of solutions proposed and the possibility of catching errors. Group thinking looks bad if success requires high-level contributions from each member. One weak member or one inadequate channel of communication can impede the group operation. Large variable groups are more likely to include one weak member.

There is no reason a priori why problems cannot be invented, or discovered in real life, which a group of three or four solves more efficiently than individual thinkers. In any case, if the individual's solution gets a score of zero while the group score is better than zero, the group is *ipso facto* more efficient—regardless of the number of man-minutes required.

PROBLEMS OF JUDGMENT

Let us begin with some statistical considerations, taking judgments of the length of a line as an example. When the average judgment of a group of judges is compared with the judgment of the typical individual, the averaging process will make the group judgment more accurate. Every judgment contains some error, and if these errors are random for the individual judges, they will be canceled in the averaging and the average will therefore have less error than the individual judgments taken separately. Suppose, for example, three judges estimate the length of a line which our best physical measurements tell us is 5 inches long. The judgments are: Mr. A, 4 inches; Mr. B, 5 inches; Mr. C, 7 inches. The average of these is 5.3 inches, so the average error is only .3 inch. The errors for the three individuals are, in order, 1, 0, and 2 inches. The average of the errors of the three individuals is therefore 1 inch. In spite of the fact that one individual was exactly right, the average of the individual errors is three times the error of the group average. When the individual errors are random, the group average will always be more accurate than the average individual, and this superiority of the group is a mathematical superiority that has nothing in particular to do with social processes. As a rule, the error of the group mean will be inversely proportional to the square root of the number of judges.

Some judges, to be sure, are better than others. They have better vision, or more experience, or they are more interested in the task. If

we could identify such people, we might find that the group average of five is more accurate than the group average of a larger number of indifferent judges. And one expert who makes small random errors could be more accurate than the average of a small group whose members make large random errors.

The errors, of course, may not be random. In the example above the group average overestimated the line by .3 inch. If we asked 400 subjects to estimate the line and this error remained, we would have to conclude that it is a constant or systematic bias that is not canceled out in the averaging process. A small group of judges free from such bias could do better than a large group all of whom are liable to a sizable systematic error *in the same direction*. And one judge might do better than a small group composed of such biased individuals.

Instead of estimating the length of each line in inches, the judges may compare one line with another and make a two-category judgment. The principles are the same, though the calculations will now be made in percentages.

Another way of doing it is to present several lines and ask the judges to rank them in order of length. A ranking that sums the rankings of a small group of judges will correlate more closely with the objectively correct rank than the ranking of one ordinary individual. And the average ranking for a large group will be more accurate than the average ranking made by a small group (Gordon, *1924*). Again, this is not a social phenomenon (Stroop, *1932*). Increasing the number of judges simply cancels random errors in the rankings, as predicted by the conventional Spearman-Brown formula—provided, of course, that the correlation with the objective rank is not zero (Preston, *1938*).

The correlation procedure eliminates one kind of systematic error. But suppose we ask our judges to rank a series of weights and we introduce irregularities in the size of these weights. Different judges will be differently influenced by the size-weight illusion. Group agreement will be reduced and so will the correctness of the group ranking (B. B. Smith, *1941*). One sophisticated individual, familiar with the size-weight illusion, might be superior to the group.

This discussion may be rephrased in the language of psychometrics. Increasing the number of judges, from one to two, or from two to fifty, will increase internal consistency or reliability if the judges added are the equals of the original judges. An increase in reliability brings along an increase in validity. But if the additional judges are

different in some way from the original judges, both reliability and validity may drop. And if there is no external criterion to which the judgments can be compared, validity (or systematic error) cannot be computed.

When the objects of judgment are complex and there is no objective criterion, e.g., when paintings are judged on an esthetic scale, different judges will judge different attributes or dimensions (see Chapter 9). The question then, in the language of factor analysis, becomes one of the importance of the general factor in the judgments as compared with more specific factors (Eysenck, *1939, 1940, 1941*).

Because of these statistical relationships we cannot prove that group judgment is better than individual judgment by comparing the mean judgment of a group with the judgments of the individuals in this group. We must compare groups working as individuals with groups of approximately the same size and composition in which some kind of social interaction has occurred.

This is exactly what Jenness (*1932*) did in the experiment described above. Two ways of expressing his results seem to be in disagreement.

1. In respect to the accuracy of group averages, groups that had the benefit of social interaction between the first judgment and the second did not improve any more than groups that made two judgments without social interaction. Group 1 got worse. Group 2 improved. Group 3 did not change.

2. In respect to the number of individuals who improved their accuracy, or, what amounts to the same measure, in respect to the average error of the individuals, social interaction did produce an improvement in accuracy.

These paradoxical results and others like them can best be understood as effects of two separate trends of group discussion: the trend toward homogeneity, and the trend of the group mean.

The trend toward homogeneity brings the individual judgments closer to the mean judgment; and when the mean judgment is close to the true value, i.e., when the systematic error is small, a necessary consequence is that the individual judgments come closer to the true value. This is the most dependable principle of superiority of group judgment over individual judgment. A special case, which has occurred in the author's classes, should be considered: if the true value lies outside the range of judgments, the average error of the individual judgments is not changed by this trend toward homogeneity. See Figure 12.7.

The trend of the mean judgment of the group toward or away from the true value after discussion is much less predictable. Chapter 9 lists some factors to be considered, e.g., prestige suggestion. The discussion in this chapter of group interaction indicates other possibilities: detection of someone's errors by someone else, reformulation of the task, pressure by one dominant individual, discovery and explanation of an illusion. Effects of this kind will not change the group mean if they are random. To improve the group mean they must shift more than

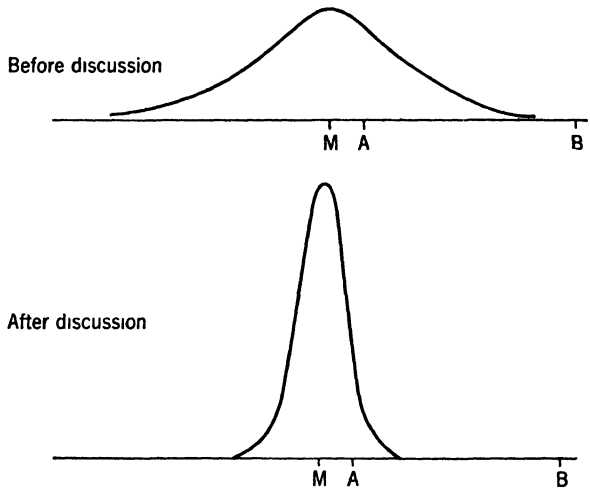


FIGURE 12.7. The effects of the trend toward homogeneity on the accuracy of judgments. The trend toward homogeneity after discussion brings the judgments closer to the mean group judgment M. If the true value, A, is close to M, the average error of the judgments will decrease. If the true value is outside the range of judgments, as at B, the average error of the judgments will not change.

half of the individual judgments *in the same direction*. Thus the trend of the group mean is a much less dependable advantage to the group judgment than the trend toward homogeneity.

These relationships are somewhat different when the group judgment is made in only two categories, but the principles are the same. In Gurnee's (*1937*) experiment on group maze learning, the groups did not discuss the pattern of the maze or plan an approach to it. They

simply voted at each point for one turn or another. The group judgment, then, was the aggregate of the individual judgments. Under these conditions small groups solved the maze with fewer errors than individuals. The groups took more time on each trial, of course, because of the voting. The same superiority of group judgments is found when a college class votes on true-false questions bearing on textbook material, general information about chemistry, and events of a three-minute movie seen just before the test (Gurnee, *1937a*).

In these experiments by Gurnee social interaction was at a minimum. All, or practically all, of the improvement could be due to the pooling of the judgments. R. L. Thorndike (*1938*) devised a method of estimating the net effect of discussion by having his subjects make their judgments individually, then again after discussion. Pooling the individual judgments yields the nonsocial majority judgment as a baseline for evaluating the net improvement due to discussion. A wide variety of true-false judgments was called for: judgments of fact, and judgments of value for which expert opinions were available and were taken as correct. Almost 1200 students in groups of four, five, or six made over 3500 decisions.

Taking all questions together, the individual judgments were 61.9 percent correct. When the judgments of the individuals in each group were pooled, 64.4 percent of these majority votes were correct. Majority votes taken after discussion were 66.2 percent correct. These differences are small but statistically significant. Even this small gain may not be the result of a social process, for we know that when college students reconsider their answers to true-false questions, a small gain in average score is frequently recorded.

All in all, the superiority of group judgment, when examined closely, seems to retreat to the vanishing point. Some of the apparent superiority is a statistical phenomenon: four judgments are better than one for the same nonsocial reason that four thermometers are better than one. The only consistent finding of significance for social psychology is the trend toward homogeneity or reduction of variance. In a secondary way this usually increases the accuracy of the individual judgments, or the size of the majority vote, but does not by itself increase the accuracy of the group judgment.

When these statistical and secondary phenomena are removed, it may still be possible to prove that group judgments are superior to

individual judgments. This has not yet been done. We need to know more of what actually happens during a group judgment. The most likely possibility for superiority of group judgment is one in which individual contributions are not added randomly, but are integrated into a group judgment that is more than the algebraic sum of the individual judgments.

An equally important topic is the group's satisfaction with a judgment or decision, its willingness to accept it and carry it through. When various groups are compared, group satisfaction with a decision is only slightly correlated with the effectiveness of the group's intellectual activities (Marquis, Guetzkow, and Heyns, *1951;* Johnson and Smith, *1953*). These two criteria of group thinking cannot be considered the same.

In many real-life discussions, especially those in which the members are vitally involved, the acceptance of the decision is much more important than the intellectual activities by which the decision was reached. Acceptance of group decisions has therefore come in for considerable study (e.g., Bradford and French, *1948;* Lewin, *1952*), and in this respect there is better evidence for the superiority of the group.

SOCIAL ORGANIZATION OF PROBLEM SOLVING

There is much to be learned from the study of individual thinkers and small groups of thinkers struggling with their isolated little problems. The story is not complete, however, without considering how society as a whole struggles with its long-time large-scale problems. Various intellectual agencies of society, such as science and art, acquire some autonomy through their professional organizations. Like the priests and the poets, however, they can never get very far away from the people if they need resources or want acclaim.

The sociology of thought cannot be thoroughly understood, of course, without some understanding of general sociological principles, the most pertinent of which are discussed in the sociology textbooks under such headings as social organization, diffusion of culture, and resistance to change.[3] The remainder of this chapter attempts merely to show how the contributions of individuals and small groups fit into the larger social enterprise, especially in our society.

[3] For an up-to-date treatment of the place of science in modern society, see Barber *(1952)*.

DIFFERENTIATION OF FUNCTIONS

As a social organization, or a whole society, becomes more complicated, its problem-solving activities become differentiated. A complex society does not solve its problems as a whole, nor do individuals as individuals solve any large-scale problems. Problem-solving functions are separated, and these different functions are performed frequently by different individuals or small groups.

The trifunctional armchair analysis of problem solving used in preceding chapters for the individual—preparation, production, judgment—can be applied without much strain to the problem solving of society. The critic, for example, is not usually a productive artist. His social function is to judge the production of the artist. The artist may prepare or formulate his own problem or he may be commissioned by a patron. The artist judges his own painting, novel, or symphony, to be sure; but from the long-range viewpoint of society it is the judgments of other artists, editors, critics, and the public that count. Table 12.4

TABLE 12.4. Differentiation of Problem-Solving Functions in Modern Society

Problem Area	Preparation: Formulating Problem proposing, Goal setting	Production: Solution proposing Development giving	Judgment: Evaluating Agreeing Disagreeing
Arts	Patron Artist	Artist	Patron Other artists Critic Editor Public
Industrial research	Administrator	Scientist	Administrator
Pure research	Scientist Textbook writer	Scientist	Other scientists Editor Textbook writer
Government	Executive Legislature	Staff Committee Commission of experts	Executive Legislature Electorate
Administration of justice	Judge Legislature Complainant	Opposing attorneys	Judge Jury

shows the typical division of functions for some familiar problem areas of modern society. Whatever utility this trifunctional analysis has when applied to the social organization of problem solving is an argument for its validity.

These different intellectual functions may be identified and assigned to different agencies of society, but some overlap will be noticed because social thinking, like individual thinking, has a wheels-within-wheels character. Critics as a rule limit their function to judging what the artist did, but some do not hesitate at times to tell the artist that he should have tried to do something different. The scientist may step out of his role as producer of solutions to act as consultant and evaluate solutions proposed by others. An artist commissioned to paint a mural for the new postoffice may stop work and change his goal—with or without the coöperation of his patron. When laws that prepare the structure of judicial contests are written, opposing attorneys are often given an opportunity to help formulate these laws. Scientists in industry are supposed to work on problems that are handed to them by the powers-that-be, but occasionally they can persuade the administrators to set a different goal. In fact the research administrator is very likely to be an ex-researcher. But by and large the differentiation of problem-solving functions is a fairly stable characteristic of complex modern societies.

SOME FACTORS THAT INFLUENCE SOCIETY'S PROBLEM SOLVING

Certainly society's efforts to solve its problems take many forms and are influenced by many dynamic and structural factors. The information assembled in preceding chapters may be extrapolated to suggest principles that slant social thinking this way or that.

PREPARATION. There is no doubt that the solution of any large problem of society or of any segment of society depends on what may be called the static aspect of society's preparation. Knowledge is accumulated in libraries, museums, and universities, and special agencies of society are designated to care for and disseminate this knowledge. The availability of information about previous successes and failures is increased by advances in communication, transportation, and general prosperity and reduced by the damages of warfare, poverty, and censorship. Cataloguers, indexers, textbook writers, and other integrators of existing knowledge play a facilitative if humble role in preparing the materials for the more dramatic functions of the creative social

units. Another static aspect of society's preparation for problem solving is the availability of trained personnel, which is influenced by the same broad factors: general prosperity and educational organization, warfare, and other social disruptions.

The dynamic aspects of society's readiness for the solution of problems exert a selective influence. Given the information and the manpower, toward which problems will they be directed? How does society formulate the problems on which it is willing to spend its resources? The answer lies primarily in the values of the society or, more precisely, the values of the segments of society that control these resources. When religious values were prominent in the *Zeitgeist* of western Europe, tremendous intellectual effort was devoted to theological problems. In the United States at present, technological progress has great prestige. A large share of the nation's intellectual resources are channeled in this direction. Problems are being formulated and solved with astounding success. Less encouraging illustrations are furnished by the problems of highway accidents and mental illness. Sporadic attempts have been directed toward these problems, but there are no large organizations with great resources—comparable to the resources devoted to the development of plastics—whose success depends on solution of these problems. Progress to date reflects the trivial nature of the attack. When problems are well formulated and the resources of information and trained personnel are available, social demands usually generate solutions. Under these conditions solutions are often brought forth simultaneously by different thinkers in different places.

The effects of social organization on preparation for problem solving may be quite specific. Research on labor-management problems, for example, is usually sponsored and financed by management. Thus the problems are formulated as problems of "labor unrest," turnover, or restriction of output. If the research were sponsored and financed by labor, we may be sure that the problems would be stated in different ways—and that different solutions would be proposed. This statement may be confidently made in ignorance of the individuals who propose the solutions.

The optimal organization of social problem solving and the role of the research administrator are new problems of modern society that have hardly been touched by empirical social science. Autobiographical

anecdotes are instructive in a preliminary way, but usually emphasize the hero and ignore the social conditions.

A new example of this relationship between social organization and scientific progress is the case of modern psychology. For many years psychological research was carried on by individual psychologists, using whatever spare time and funds they could extract from their academic situations. They formulated their own problems, influenced of course by their own motives, the values of the middle class that most of them belonged to, and their beliefs about the kind of research that would be accepted by other psychologists (see Murphy, *1942*). Financial resources were slight. Now that psychologists have demonstrated their ability to solve practical problems of many kinds, sponsorship and large research funds are coming from many sections of society. The impact of this largess on the progress of psychological science will not be clear for some time, but psychologists in the middle of this stream are debating "the effect of military contract research on the field of psychology" (see Lanier, Bray, Wilson, and Darley, *1952*), and some of the opportunities and limitations are already apparent in this symposium.

Production of Ideas. When a social organization such as a corporation, a governmental or military unit, a club, or a church identifies a problem, it frequently turns it over to a regularly designated research-and-development unit, or an *ad hoc* committee of "idea men." Their function of producing ideas as possible solutions is recognized as different from routine administration, selling, teaching, or public relations. It is usually carried on by small groups or individuals.

Little is known of the social conditions that facilitate this productive activity. One would expect that a minimum of social security and freedom from other responsibilities is necessary, but there is no direct evidence. The novelist must eat before he can write, but history has shown that he does not have to eat very much. Surely some social organizations facilitate productivity of thought more than others, and some hypotheses may be drawn from the research on small groups summarized above. Communication with other thinkers engaged in similar problems would seem to be helpful. Schools of art and of science generate enthusiasm and fluency of output. Assurance that one's ideas will be seriously received should be encouraging. For some kinds of problem, as pointed out above, sheer number of thinkers at work must be a factor. The testimony of successful research organizers is instruc-

tive but is not proof—any more than Helmholtz's testimony about his own creative thinking is proof that his description is accurate.

Social Evaluation of the Solution. Several ways of formulating a problem and several proposals for its solution are usually under discussion more or less seriously for some time. During this period some individuals may make up their minds, some may suspend judgment, and some may ignore the matter. Typically the discussion comes to a climax and opposing social forces converge when a definite decision is required at a designated time. In our society these decisions are made by the electorate; by small groups, such as boards of directors, juries, and committees; or by individuals to whom the power of decision has been given by appointment, election, or custom. Preparation and production are often carried on backstage, but important judgments are usually delivered in our society, and most others, with suitable ritual. Ceremonial announcements, signatures, witnesses, the solemn filing in and out of impressively tailored dignitaries, all attest the seriousness of the decision and give official notice that it shall not be easily undone.

All the principles mentioned earlier, governing the judgments of the individual and of the small group, enter into these social decisions. The important additional factor is that when the decision is a socially significant one, it measures the relative strength of opposing social forces.

REFERENCES

Bales, R. F., 1951. *Interaction process analysis,* Cambridge, Mass., Addison-Wesley Publishing Co.

Bales, R. F., 1952. Some uniformities of behavior in small social systems, in *Readings in social psychology,* rev. ed., by G. E. Swanson, T. M. Newcomb, and E. L. Hartley, New York, Henry Holt & Co.

Bales, R. F., and Strodtbeck, F. L., 1951. Phases in group problem-solving, *J. Abnorm. & Social Psychol., 46*:485–495.

Barber, B., 1952. *Science and the social order,* London, George Allen & Unwin.

Bovard, E. W., Jr., 1951. Group structure and perception, *J. Abnorm. & Social Psychol., 46*:398–405.

Bradford, L. P., and French, J. R. P., Jr., eds., 1948. The dynamics of the discussion group. *J. Social Issues, 4*:2–75.

Carter, L., Haythorn, W., Meirowitz, B., and Lanzetta, J., 1951. The

relation of categorizations and ratings in the observation of group behavior, *Human Relations, 4*:239–254.

Cattell, R. B., and Wispe, L. G., 1948. The dimensions of syntality in small groups, *J. Social Psychol., 28*:57–78.

Dashiell, J. F., 1935. Experimental studies of the influence of social situations on the behavior of individual human adults, in *Handbook of social psychology,* ed. C. Murchison, Worcester, Mass., Clark University Press.

Deutsch, M., 1949. An experimental study of the effects of cooperation and competition upon group process, *Human Relations, 2*:199–231.

Eysenck, H. J., 1939. The validity of judgments as a function of the number of judges, *J. Exper. Psychol., 25*:650–654.

Eysenck, H. J., 1940. The general factor in aesthetic judgments, *Brit. J. Psychol., 31*:94–102.

Eysenck, H. J., 1941. The validity and reliability of group judgments, *J. Exper. Psychol., 29*:427–434.

Fouriezos, N. T., Hutt, M. L., and Guetzkow, H., 1950. Measurement of self-oriented needs in discussion groups, *J. Abnorm. & Social Psychol., 45*:682–690.

Gordon, K., 1924. Group judgments in the field of lifted weights, *J. Exper. Psychol., 7*:398–400.

Guetzkow, H., 1950. Unitizing and categorizing problems in coding qualitative data, *J. Clin. Psychol., 6*:47–58.

Gurnee, H., 1937a. Maze learning in the collective situation, *J. Psychol., 3*:437–443.

Gurnee, H., 1937b. A comparison of collective and individual judgments of fact, *J. Exper. Psychol., 21*:106–112.

Heise, G. A., and Miller, G. A., 1951. Problem solving by small groups using various communication nets, *J. Abnorm. & Social Psychol., 46:* 327–335.

Hemphill, J. K., 1949. *Situational factors in leadership,* Columbus, Ohio State University.

Heyns, R. W., 1948. Functional analysis of group problem-solving techniques, unpublished report of Conference Research project, Ann Arbor, University of Michigan.

Husband, R. W., 1940. Cooperative versus solitary problem solution, *J. Social Psychol., 11*:405–409.

Jenness, A., 1932. The role of discussion in changing opinion regarding a matter of fact, *J. Abnorm. & Social Psychol., 27*:279–296.

Johnson, D. M., and Smith, H. C., 1953. *Democratic leadership in the college classroom,* Psychol. Monogr. 67, No. 361.

Lanier, L. H., Bray, C. W., Wilson, J. T., and Darley, J. G., 1952. The

effect of military contract research on the field of psychology, *Am. Psychol.*, *7*:707–721.

Leavitt, H. J., 1951. Some effects of certain communication patterns on group performance, *J. Abnorm. & Social Psychol.*, *46*:38–50.

Lewin, K., 1952. Group decision and social change, in *Readings in social psychology*, rev. ed., by G. E. Swanson, T. M. Newcomb, and E. L. Hartley. New York, Henry Holt & Co.

McCurdy, H. G., and Lambert, W. E., 1952. The efficiency of small human groups in the solution of problems requiring genuine cooperation, *J. Personality*, *20*:478–494.

Maier, N. R. F., 1950. The quality of group decisions as influenced by the discussion leader, *Human Relations*, *3*:155–174.

Marquis, D. G., Guetzkow, H., and Heyns, R. W., 1951. A social psychological study of the decision-making conference, in *Groups, leadership and men*, ed. H. Guetzkow. Pittsburgh, Carnegie Press.

Mintz, A., 1951. Non-adaptive group behavior, *J. Abnorm. & Social Psychol.*, *46*:150–159.

Murphy, G., 1942. Psychology and the post-war world, *Psychol. Rev.*, *49*: 298–318.

Murphy, G., Murphy, L. B., and Newcomb, T. M., 1937. *Experimental social psychology*, New York, Harper & Brothers.

Preston, M. G., 1938. Note on the reliability and validity of group judgment, *J. Exper. Psychol.*, *22*:462–471.

Shaw, M. E., 1932. A comparison of individuals and small groups in the rational solution of complex problems, *Am. J. Psychol.*, *44*:491–504.

Sherif, M., 1935. *A study of some social factors in perception*, Arch. Psychol., No. 24.

Smith, B. B., 1941. The validity and reliability of group judgments, *J. Exper. Psychol.*, *29*:420–426.

Smith, H. C., and Dunbar, D. S., 1951. The personality and achievement of the classroom participant, *J. Educ. Psychol.*, *42*:65–84.

South, E. B., 1927. Some psychological aspects of committtee work, *J. Appl. Psychol.*, *11*:348–368, 437–464.

Stroop, J. R., 1932. Is the judgment of the group better than that of the average member of the group? *J. Exper. Psychol.*, *15*:550–562.

Taylor, D. W., and Faust, W. L., 1952. Twenty questions: efficiency in problem solving as a function of size of group, *J. Exper. Psychol.*, *44*: 360–367.

Thorndike, R. L., 1938. The effect of discussion upon the correctness of group decisions, when the factor of majority influence is allowed for, *J. Social Psychol.*, *9*:343–362.

Watson, G. B., 1928. Do groups think more efficiently than individuals? *J. Abnorm. & Social Psychol., 23*:328–336.

Weld, H. P., and Danzig, E. R., 1940. A study of the way in which a verdict is reached by a jury, *Am. J. Psychol., 53*:518–536.

Weld, H. P., and Roff, M., 1938. A study in the formation of opinion based upon legal evidence, *Am. J. Psychol., 51*:609–628.

Wright, M. E., 1942. Constructiveness of play as affected by group organization and frustration, *Character and Person, 11*:40–49.

Chapter 13

The Improvement of Thinking

Compose with fury and correct with phlegm.

IT MUST be admitted at the outset that little is known about the improvement of thought. But this problem is so important that it must be taken seriously; even small improvements and suggestions for improvement are worth talking about.

Nothing is so practical as a good theory, and no test of a theory is so convincing as its application to practical problems outside the field in which the theory was constructed. The principles formulated in earlier chapters suggest many recommendations for the improvement of thinking, and the testing of these recommendations would test the adequacy of the formulation.

The very statement of the question is not simple. It is not a question of describing the conditions under which thinking is done effectively, nor of identifying those who think well. These questions have been considered in preceding chapters. The present question is rather one of practicality. What steps can be taken, by the thinker himself or by someone else, and at what times and places, that will improve the outcome?

GENERAL RECOMMENDATIONS

Recommendations for improvement should refer to specifiable controllable actions. It is useless, for example, to tell someone to "keep an open mind" unless we tell him how it is done. Hanging a neatly framed signed on the wall that reads THINK is about as efficacious as telling someone not to worry. Some feasible recommendations may be listed, however, as examples of what can be done.

Ingestion: The thinker can take a drug or a vitamin or a brain food. Or he can avoid certain foods.

Working conditions: The thinker can open a window. He can soundproof the walls. He can arrange his books or his paint brushes or his calculating machine conveniently.

Scheduling: The thinker can plan his heavy thinking for a certain preferred time of day. He can work in 45-minute periods. If he can identify certain intellectual operations, he can schedule them in optimal sequences. He can alternate productive activities with critical activities.

Learning: The thinker can prepare by mastering pertinent facts, concepts, and principles. He can study the common errors and successful solutions of other thinkers.

Thought control: The thinker can adopt a set to attend to some things and ignore others. He can do this by writing down a rule or guide to follow. Or he may repeat certain sentences to himself. He can cue one activity to another; that is, when he starts to quit, he can remind himself of rewards for achievement.

Any of these acts can be performed by the thinker himself or by his manager, teacher, or spouse. Obviously the effectiveness of any technique cannot be verified unless instructions for the operation can be communicated to others.

POINT OF APPLICATION

Improvement can be attempted in a global unanalytical way simply by changing the conditions of work and observing the results. Many people drink a cup of coffee to improve their thinking without considering the nature or the locus of its effect. It is reasonable to suppose that anything that improves the efficiency of the organism will improve thought and judgment. All that is required is to specify the antecedent conditions and check on the end results.

Much better results can be expected if attempts to improve thinking are based on a reasonably good analysis of thinking. Several processes or functions are assumed. The adequacy of the result of thought depends on the adequacy of all contributory functions in the system. If the adequacy of each function can be assessed, the most likely sources of failure can be located, and an effort at improvement can be applied where it will do the most good. Effort applied at any point in the system may be helpful, but it is better to diagnose the difficulty and apply the effort at the point of greatest difficulty. When Burack (*1950*),

for example, studied the efficacy of various methods of attack on three kinds of problem, he discovered that the function that caused difficulty in one problem might be very easy in another and, in particular, that formulation of the problem was so obvious that there were no difficulties at this point in any problem. Bloom and Broder (*1950*) observed similar variations.

Utilization of this principle requires that the analysis be approximately correct and that some method be available for locating the difficulties. There are at least four possibilities:

1. The nature of the final result, especially if there are errors in the solution, may permit inferences about the locus of the errors.

2. If the functions can be examined separately, direct observation of the defective functions may be possible.

3. The source of an error may be located *ex post facto* by correcting it. If the conditions of preparation are improved, for example, and the final result improves, there must have been something wrong with the preparation.

4. A rather complicated indirect way of finding the source of an error is by way of individual differences. Suppose A and B make errors in solving a problem, while C and D are successful. Let us test all four individuals on another problem artificially simplified so that errors in each function can be observed. Or we can test them on artificial problems that depend largely on only one function. If A and B are inadequate on one particular function in these artificial problems, while C and D are successful, we may infer that A and B are inadequate in this function when it is embedded in a complicated problem.

These ways of locating the difficulty in thinking have been stated in respect to errors, as if each function and the end result were either right or wrong. This is an oversimplification, to be sure. Some criterion of goodness is implied in the very notion of improvement, but the contribution of an intellectual function to the whole enterprise may take several forms. The adequacy of the function may be of an either-or, right-wrong nature. If the formulation of the problem, for example, is above a threshold of suitability, no improvement is necessary or possible. Or, to take a very simple example, if the problem is a mechanical one, the illumination must be sufficient that the thinker can see the materials at hand. Improving the illumination beyond this point will not be helpful.

The adequacy of other functions may be continuous. Any improve-

ment in conditions that affect productivity, for example, would appear to improve the end product of thought—without limit.

Some other functions may be characterized by an optimum, which is approached asymptotically. Large departures from this optimum do more harm to the whole enterprise than small departures. Judgment of the solutions produced may operate in this way. Too much criticism is as bad as too little. Efforts toward improvement of such functions should be directed toward an optimum or balance—as implied in the phrase good judgment.

If all this is true, the overall strategy for improving thinking, or any other complex system of operations, would consist of (1) analysing the system of intellectual operations, (2) diagnosing the weaknesses, (3) applying efforts at improvement to the points of greatest weakness.

TRANSFER OF IMPROVEMENT

Obviously techniques for solving arithmetic problems can best be improved by working arithmetic problems. Techniques for solving economic problems can best be improved by working economic problems. But one never knows whether the problems he will face tomorrow will be economic or arithmetic or psychological. So it is natural to seek some techniques of improvement that the thinker can transfer to all kinds of problems. The results to date have not been encouraging.

Hill (*1936*) investigated this question of transfer by training four groups of students on different kinds of reasoning problems: arithmetic reasoning, logical reasoning, interpretation of fables, and good judgment in problem situations. Each group improved more than a control group in respect to the problems on which they were trained, but there was very little transfer to the other three kinds. Wesman's (*1945*) results were similar.

It is more encouraging to remember that two problems may not be so different psychologically as they first appear, especially if both are complicated and abstract. Likewise, two problem-solving techniques may be more alike than they first appear. Many problems depend on mastery of abstract verbal relationships, and skill in solving these can be improved by training in interpretation and discussion of the printed word (Simpson, *1923;* Barlow, *1937*) and in logical analysis (White, *1936*). Training in mathematics transfers to other problems when the

mathematics is taught by special methods which make the similarities explicit (Winch, *1923;* Ulmer, *1939*).

Improvement in handling abstract verbal and numerical material will transfer to a wide variety of problems because many problems are either formulated in abstract terms or are solved by the aid of instrumental verbal and numerical patterns. Furthermore, the abstractness of problems is known to be a frequent source of difficulty. Improvement in handling facts and spatial relationships probably transfers less widely.

Any improvement that can be produced in the dynamics of thought could transfer to almost all problems. Fortunately Updegraff and Keister (*1937*) have shown that when children begin on easy problems and progress to harder ones, they approach the new problems with confidence and persistence. Glaser (*1941*) found that of all the aspects of thinking he tried to train, an attitude of thoughtfulness was the most susceptible to improvement.

An interesting set of teaching materials for children in kindergarten and the first grade has been prepared by T. G. Thurstone (*1949*) and called the Learning to Think Series. The books consist of many problems intended to give training in eight abilities identified by factor analysis. In fact the problems are the same as the items used as tests of these factors. The advantage claimed for this series is the inclusion of training in abilities omitted in conventional teaching materials. For example, the *Green Book,* intended for the second half of the first grade, has six types of reasoning problems: classification of pictured objects, association by induction, part-whole relationships, discovery and use of a rule in stringing beads, grouping pictures according to a principle, and rearranging pictures to make a story. The series "has been constructed to give young children direct training in the mental skills that will be most helpful to their later success in school and life."

It would be difficult, of course, to prove that the intellectual abilities developed by this set of materials will transfer more widely than the abilities developed by other sets of materials. But starting children on abstract problems at an early age and in an interesting manner has important dynamic implications. If the children develop a favorable general attitude toward abstract material, this attitude could carry over to a wide variety of problem situations in school and out.

The general recommendation about transfer, then, is this: Don't expect very much. If we know what specific problems someone will have

to solve, we can train him to solve problems of this kind, usually with good success. If we do not know what kind of problems his life will lead him into, we should train him for a variety of problems, that is, for thinking in general. Training for thinking in general will transfer to all kinds of problems, though the amount of improvement in any one may be small.

PREPARATION

Testimony of outstanding thinkers, records of achievement, and laboratory experiments all agree that the preparation for thinking is very important. Since this function is often skipped over hastily by unsophisticated thinkers, this may be the point at which efforts at improvement can be applied most successfully.

PREVIOUS LEARNING

Certainly a supply of pertinent facts, concepts, and principles is necessary for successful thinking. We do not always know just what information will be pertinent, for the crucial fact or principle for the solution of an economic problem may be in psychology or politics or law rather than in economics proper, but the recommendation is clear in principle, and basic.

Some ways of acquiring information for later use in thinking must be better than others. Learning is not altogether distinct from thinking (see Chapter 5), so it is a reasonable hypothesis that the best kind of learning in preparation for thinking is learning that is most similar to thinking.

A teaching experiment by Kilgore (*1941*) is in effect a test of this hypothesis. He stressed principles of physics in high-school classes with the aid of special teaching materials directed toward those principles that are useful in practical life outside of school. When these experimental classes took a final examination that required solving physics problems, it turned out that they had made much larger gains than control classes taught by conventional methods. The experimental classes were not superior to the control classes on an ordinary physics test that did not require so much problem solving.

A recent experiment by Szekely (*1950*) helps to clarify the relations between learning and subsequent thinking. His 40 subjects got four typewritten pages of principles of mechanics, including momentum of

rotation. Half of the subjects learned by the traditional method of lecture and demonstration, common in European schools; the other half learned in a more thoughtful way. A puzzling phenomenon was demonstrated, and the students were told that they could not predict what happened; but if they would study the four pages, they would learn how to predict it. After a few days all were given the problem of comparing a hollow sphere with a solid sphere of equal size and weight to discover which one was hollow. The solution depended on application of one of the principles studied in the previous training period. Of the 20 who studied by the traditional method only 4 solved the problem; of the 20 who studied by the more thoughtful method 13 solved it.

A later checkup on memory for what had been learned was very revealing. Of the thoughtful learners only four were able to repeat verbatim the definition of momentum, and none of these had solved the critical problem. Of the traditional group eight were able to repeat the definition correctly, and only two of these had solved the problem. Apparently, in the group that learned by the traditional method, learning and problem solving were different functions. Szekely concluded that the productive application of a principle and its verbatim reproduction appear to be mutually exclusive.

Introspective reports were instructive in a negative way. The subjects in the thoughtful group who were successful could not tell from what sources their insights had developed. The subjects in the traditional group who were successful were more explicit and less intuitive about the source of their insights.[1]

The best recommendation about learning for problem solving is to learn those facts and principles that will be used later. The best recommendation about techniques of learning is to make the learning similar to problem solving.

FORMULATING THE PROBLEM

A collection of facts and principles is not enough. The pertinent data must be organized into a pattern or schema with the gap for the solu-

[1] Suggestions for learning procedures that will facilitate later problem solving may be found in standard treatments on educational psychology. For a recent summary of up-to-date recommendations, see the *49th Yearbook* of the National Society for the Study of Education, Part I (N.S.S.E., 1950). For the special case of arithmetic, see the *50th Yearbook*, Part II (N.S.S.E., 1951). For the use of symbols in thinking, see Johnson (*1953*).

tion clearly indicated. This gives a starting place and guide for the production of promising solutions.

The common metaphor which makes facts, images, concepts, principles, etc., the tools of thought is misleading. These are the raw materials of thought. The tool is the schema which, like a template, shapes the raw material into a previewed solution.

It is wise to be quite explicit about formulating the problem, clarifying what is put into the schema and what is left out. Arithmetic teachers often recommend writing down what is given and what is required before actually attempting to produce solutions. A common error to be avoided is the tendency to pick up some aspects of the preparation casually from the environment or atmosphere in which the thinker works. All aspects of the preparation should be examined critically. Any graphical device, such as diagraming or outlining the problem at hand, should be helpful.

The problem may be formulated or narrowed down from either end. We may begin with what is given and work toward what is required, or we may begin with what is required and work toward what is given.

It is wise to be quite explicit about intellectual operations also, if we wish to promote transfer of improvement. If errors are frequent in the formulation stage, the thinker should identify formulation and differentiate this operation from production. Simplified assignments, like discrimination, word association, and simple problems in which it is easy to separate preparation from production, should help the thinker examine the preparation process critically. What we are aiming at really is a kind of thought control, or at least control of the preparatory set.

Instructions from someone else may not be the best technique of improvement. Practice in formulating and reformulating problems may be more effective. Some exercises may be suggested for this purpose—without proof of their efficacy.

Assemble tools and materials, then stop and examine them critically before beginning work.

Formulate a mathematical problem algebraically, then graphically, then verbally. Reformulate it in each of the three ways.

Begin with an assemblage of data relative to a problem, then eliminate what is not pertinent. If pertinent data are lacking, describe the kind of data that are necessary.

Begin with a complex social problem, formulate it quickly, and begin producing solutions. Then go back and list assumptions of first formulation.

Let several thinkers independently formulate an approach to a psychological problem. Exchange formulations and examine each critically.

The shelves of the public libraries hold many books, written from several points of view, that tell their readers how to think effectively. Most of them contain at least a few worth-while suggestions. For high-school youth, Seashore and Van Dusen (*1950*) have prepared a pamphlet, *How to Solve Your Problems,* that describes a scheme for attacking personal problems by listing obstacles, assets, possible solutions, and probable results in a systematic way. All such schemes have the value of aiding explicit formulation of the problem.

Formulation of some problems under some conditions may be so obvious that no errors are made in this operation. This reëmphasizes the necessity for diagnosis of the major difficulty. And of course there are differences between individual thinkers. Cautious thinkers, who spend lots of time in preparation before getting under way, do not need this practice. We should identify the thinkers who jump into production without due preparation and give them the special practice in preparation procedures.

CONTROL OF EMOTION

It is fashionable to contrast thought with emotion, as if the best thinking takes place in an aseptic, dispassionate atmosphere. This is an oversimplification. It would be more accurate to speak of an optimum of emotional involvement. There is no problem without motivation. The difference between solving a problem and failure is often the persistence that comes from genuine motivation. An intrinsic interest in the problem for its own sake is of course preferable to an extrinsic motivation toward the results alone. Attempts to improve thinking should avoid the ideal of a cold, impersonal machine, and should try to arrange conditions so that the motivation is maintained.

The optimum is exceeded when the motivation reaches emotional heights. We cannot locate the optimum degree of arousal precisely—it probably differs for different organisms—but in general the production of new ideas requires some degree of enthusiasm, while

judgment of these ideas is improved by a calm critical detachment. Figure 13.1 attempts to show these optima schematically.

Better than designating optimal emotional involvement a priori is to watch for the deleterious effects of emotion: repression of pertinent data or their implications, and loss of flexibility in reorganizing the data or in shifting the set.

But how does one control his emotions, or get others to control theirs?

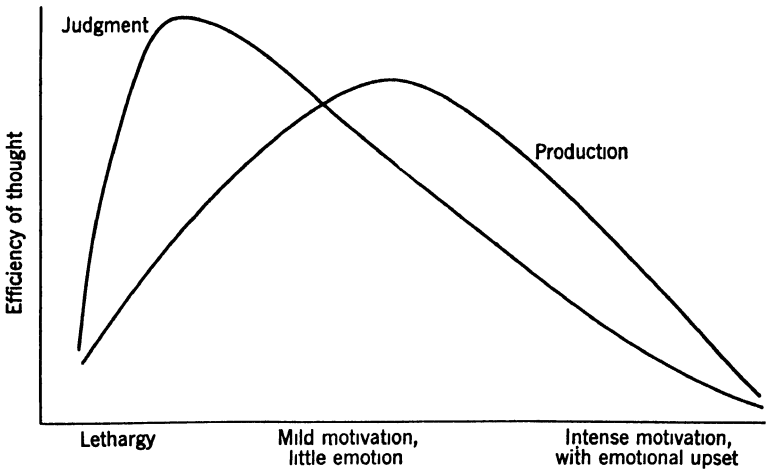


FIGURE 13.1. Hypothetical curves schematizing the effects of motivation and emotional involvement on two thought processes.

One standard method is to ventilate the difficulty. For good thinking it is probably not necessary to clear up all the quirks of one's personality, but it should help to open up the biases that are pertinent to the problem at hand. It is not at all uncommon for a serious thinker to admit a prejudice and try to discount it. The danger of overcorrection must also be considered.

For those who are willing to go the limit and undertake psychotherapy, the improvement in thinking may be considerable. Magaret (*1950*) has listed the results of successful psychotherapy as changed reaction sensitivities, increased flexibility of behavior, and success in handling new situations. The basis of these improvements in thinking is the formation of learning sets. The patient learns how to handle

interpersonal problems in the therapist's presence and generalizes this improvement to other interpersonal situations.

There are many people with personality problems who solve other problems quite well. Apparently they are able to divorce themselves at times from their emotional difficulties. It is possible even that some creative activity is promoted by a flight away from the annoyances of reality.

The recommendation that follows is that the ordinary thinker with the ordinary assortment of personal difficulties need not wait until his difficulties are completely eradicated, but should try to isolate himself intermittently from his difficulties. Spatial separation helps. Some people with troubles at home schedule their important decisions for the office. Some do the reverse. The general rule is to arrange conditions so that emotional and thoughtful episodes occur at different times.

Most of these recommendations are negative, and positive recommendations are usually more effective. If one can develop a genuine interest in the problem at hand, he may be able to bypass, or at least overcome, his prejudices, his overcorrections, and his inhibitions. When an artist gets absorbed in his painting, he loses interest in himself. When a physician becomes concerned about his patient's health, he pays more attention to the color of his urine than to the color of his skin. Professional social scientists are less liable to emotional thinking about social problems than the laity, not because they are nicer people, but because they have a professional interest in these social problems.

One way to develop this interest in the problem *per se* is suggested by the experiments of Updegraff and Keister (see page 485) and Lefford (see page 297). Schedule easy problems first, without emotional involvement. Then schedule problems in which emotional distortion is possible but is easily recognized and discounted. Then increase the possibilities of emotionality. Surely this sequence would be better than the reverse. College professors who introduce graduate students to research have often noted that if the first problem undertaken is near or beyond the student's capacity, the anxiety generated may lead to failure—even to failure on a subsequent easier problem.

IMPROVEMENT OF PRODUCTION

Preparation for thought should be deliberate, critical, serious. Production is a different process. It should be enthusiastic, fluent, provoca-

tive. Of course it is difficult for anyone to control his mood, but it is not altogether impossible. A warm-up period may be helpful. Playing games just before tackling a hard problem or at times when progress seems to be blocked is worth trying. President Lincoln used to tell dialect stories when his advisers could not settle the pressing problems of state. Practice in free association may have a loosening effect.

The danger to be avoided is rigidity or loss of flexibility. Now telling someone to be flexible is probably a waste of time, but the phenomenon itself can be identified. When the thinker produces the same or similar ideas repeatedly, it is easy to see that he is in a rut. When this occurs, the thinker can drop the problem temporarily and go for a walk or at least do something different. Abandoning the problem when the production curve levels off seems to be the best technique for eliminating interferences and allowing incubation to occur.

It is possible that coffee, Benzedrine, alcohol, or a cold shower could promote productivity, but individual variations in the effects of these are likely to be large. A better argument could be made for the benefits of a good night's sleep. If time is not important and the solutions produced will be evaluated under sober conditions, almost any radical manipulation of the thinker's bodily condition is worth trying. At present no specifics are known and the best rule to follow is simply that any improvement in the overall functioning of the organism may include an improvement in intellectual functioning.

Many people working on a problem may develop interferences at the same point in their progress. In these cases special hints delivered at the critical time will be helpful. Or they can be instructed in advance to shift set when they reach this critical point.

Another danger to be avoided is failure to exhaust the possibilities of an approach. Flightiness is as bad as rigidity. It is not wise to retool while production is going on. One remedy is to follow a systematic plan of production. In producing words to fit a certain schema, for example, the thinker can proceed alphabetically, or by naming all the words in one class after another. He should be explicit about abandoning one approach for another, perhaps making a note that this approach is or is not worth later reconsideration. The optimal productive activity is not halfway between rigidity and flightiness, but rather an explicit decision to continue or to change one's activity and a willingness to return if necessary. Production should be a gay, lighthearted process, but a decision to abandon a plan of production should

be made seriously. An enforced pause at a critical time may be very helpful.

Because children seem to enjoy using their imagination and may actually be more productive relative to their limited experience than adults, it has often been supposed that adults have learned the bad habit of inhibiting their productivity through years of being critical. Suppose we ask people to forego criticism temporarily and deliberately exercise their imaginations for ten minutes each day. Some suggestions for exercising imagination may be listed for illustration.

Draw the most peculiar animal you can.
Make the silliest face you can.
Write nonsense verse.
Make weird noises.
Make an ugly Halloween mask.
Make up original names for a baby elephant, for a baby flea, for a baby automobile.
Rearrange the days of the week, the months of the year.
Invent silly imaginative tasks like the above.

It is possible that practice on such imaginative assignments carried out in an atmosphere of freedom from criticism and praise for originality would actually release creative potentialities. Proof of this hypothesis would necessitate reliable tests of creative imagination, to be given before and after such practice.

A common source of difficulty, when the problem is complex, is the amount of material to be organized. There is too much to attend to at once. Several tricks have been recommended for the production of a pattern that will organize the material and make it more manageable.

Sometimes simple counting will suffice. Many failures of planning can be avoided if someone goes to the trouble to count the number of people involved, the number of chairs needed, and so on. Arguments about the differences between men and women, between Protestants and Catholics, can often be settled simply by counting known facts.

When the material of the problem is more heterogeneous, the problem may often be broken into subproblems that are easier to organize. Or the data may be organized by means of a graph, map, or check list. It is standard practice in business, where the number of pertinent facts is apt to be large, to solve problems with the aid of balance sheets, summaries of operations, prospectuses, etc. This is a good model to

follow in solving personal problems. Any device that groups the data or organizes the facts so that the number of items is not beyond the span of attention should be helpful if the device does not distort the data. One must accept these visual aids only tentatively and be willing to represent the facts at hand in some other way that may fit the present purpose better.

Another technique for handling a complicated mass of data that is often used casually and should be used explicitly is the formation and testing of a hypothesis. Consider as many facts of the problem as possible and make a guess from these. Then apply this guess or hypothesis to the remainder of the facts. If it does not fit, use part of the facts to form another hypothesis, and try this out on the remainder of the facts. It may then be apparent that more facts are necessary.

In general, when the number of things to attend to is large, it is better not to try to grasp all at once and risk missing something important. Separate the facts into groups and consider them serially.

Another source of difficulty in dealing with complex problems is the abstractness of the patterns to be produced and manipulated. The initial contact with the material is important because errors in the use of an abstract pattern are not easily recognized, hence one can persist in the use of an inadequate pattern for some time. The usual way of improving operations on abstract material is to make the material more concrete, by variegated illustrations, by diagrams, outlines, and models.

For solving problems in special fields, e.g., mathematics and personnel relations, good teachers and expert practitioners usually know some special techniques that they can teach to beginners: how to arrange one's work so that the essential relations are apparent, what errors to watch for, likely sources of hypotheses, and so on.

IMPROVEMENT OF JUDGMENT

The preparation for judgment can be more specific than the preparation for productive thought. The first step is to get set for the stimulus attribute to be judged. If this stimulus attribute is abstract or concealed, it is well to schedule a few easy judgments first, in which the thing to be judged stands out prominently. For example, if applicants for a job are to be compared in respect to Trait X, applicants high and low in Trait X can be compared first so that this trait can be isolated from the other differences between the applicants. Agree-

ment between judges can be increased by having them work independently, then get together and check points of agreement and disagreement, and then work independently again. If this procedure is followed, their agreement in respect to Trait X may increase even when neither judge can define this abstract thing clearly. As they check disagreements they will probably also come closer together in their use of the response categories on the scale of judgment.

Improvement in the reliability of judgment, that is, of agreement between judges, usually improves the validity of the judgments, but it is better, if possible, to work on improvement of validity directly. An ideal task for the improvement of judgment is that of the center fielder. When the batter knocks up a high fly, all fielders try to judge its direction. The center fielder and the left fielder could get together and check on their agreement, but it is not necessary because the ball will soon land and then they can check on the ultimate validity of their judgments. Improvement, up to a limit, is almost certain because errors of judgment of the ball in the air are immediately apparent when it lands. The effects of practice on perceptual judgments have recently been summarized by Gibson (*1953*).

Psychologists often estimate the intelligence of their clients from casual conversation before they administer an objective test of intelligence. These casual judgments usually improve because they are soon compared with the objective test scores. To improve judgments of applicants for a job it is wise to keep records of the initial judgments in writing, then check these later with actual performance on the job.

The general recommendation for training judgment is to arrange conditions so that a series of judgments can be compared with some better measures. These comparisons, good and bad, should be reviewed just before another series of judgments is made.

The results of training will probably be stepped up by identification of common errors, as listed in preceding pages. The possibility of a halo effect can be minimized by judging one factor at a time and perhaps by a mere warning about halo. Instructions to ignore some factors, e.g., physical appearance when judging character, should be helpful. The danger of missing something significant when making judgments of complex matters can be minimized by a check list or printed form for assembling the data, such as bankers use when judging applications for a loan. Poor Richard advocated, as an aid in

deciding about a course of action, writing the advantages down on one side of a page and the disadvantages on the other.

Flexibility is necessary in judgment as in production. The data of complex judgments can be examined from several points of view and in several contexts. The final judgment can be expressed in several ways. But which one is best? No general statement is tenable, not even the statement that a broad context of judgment is better than a narrow one. It is safe, however, to recommend shifting contexts and points of view before coming to the final decision. Attention is usually directed to the object of judgment itself; the context is taken for granted without examination. But instruction and progressive practice in shifting context may promote flexibility. As an illustration, we might ask for judgments of a series of men in respect to all men, then in respect to men who have applied for a certain job, then in respect to all women. Such practice may improve judgment by encouraging the judges to take the context of judgment as seriously as they take the object of judgment.

The form in which the judgment is expressed is also subject to improvement. The most primitive judgments are made in two categories, such as Yes and No, or Right and Wrong. But the first system of response categories one thinks of may not be the best. We can ask our judges to judge the matter under discussion Right or Wrong, then ask them to invent another system of categories that might be more suitable. Just as thinkers can learn to be critical of the initial formulation of the problem, they can learn to be critical of the initial form in which the conclusion is expressed, and can shift the system of categories until they produce one that fits the data and the purpose of the judgment. In general the judging process is most effective when the judge and those to whom the judgment is communicated have the same understanding of the meaning of the categories of response.

When the thinker revises the response categories to permit two or more judgments instead of one, the accuracy of the judgments may improve. It may be very difficult to judge an issue Right or Wrong but easier to judge it Right in one context and Wrong in another. Such fence-straddling annoys rigid people, but it is commonly done by good flexible judges—and logicians.

Examination of all the objects of judgment, from all promising points of view, within all plausible contexts certainly takes time and persistent effort. So does critical examination of the system of response

categories. Hence conditions that encourage delay and careful consideration of the matter at hand should improve judgment. Hasty judgment is almost the opposite of good judgment.

Time, however, is not always unlimited. Some decisions must be made under pressure. In these circumstances wise men try to prepare in advance for the decisions. It is often possible to anticipate the matters that may arise, the facts that may be produced, the pros and cons, and do much of the deliberation long before the moment of decision arrives. Snap judgment is usually not good judgment. If it is good judgment, it has probably been preceded by deliberate judgment.

CRITICAL THINKING

The use of logical reasoning and the avoidance of common fallacies in judgment may be called critical thinking. Critical thinking is often listed among the aims of modern education, and several educators have made attempts from time to time to devise means for improving this skill. Training in logic is helpful, of course (White, *1936;* Anderson, Marcham, and Dunn, *1944;* Reiner, *1947;* Melzer, *1949*). Training in verbal skills, as in the interpretation of fables, improves verbal reasoning (Simpson, *1923;* Barlow, *1937*). Modern methods of education in New York City schools seem to be slightly (but not significantly) superior to conventional methods in improving ability to explain and apply facts about social affairs (Wrightstone, *1938,* page 88). Techniques of propaganda and their detection can be taught (Biddle, *1932;* Jewett, *1940*). Such knowledge does not necessarily make one more resistant to propaganda (Osborne, *1939*).

Glaser's (*1941*) attempt to improve critical thinking is one of the most thorough. He had high-school students taught some of the essentials of logic and the psychology of thinking, with practice in the recognition of errors and critical evaluation of newspaper articles. After ten weeks the four experimental classes were significantly superior to the control classes on the Watson-Glaser Tests of Critical Thinking. Critical thinking scores and intelligence test scores correlated .46, but the training did not produce significant gains in the latter.

Encouraging gains in critical thinking have been reported as a result of high-school training in debating (Howell, *1943;* Brembeck, *1949*). Brembeck's careful study analyzed the improvement of 202 students in 11 different schools, taking courses in argumentation from teachers of different educational backgrounds. Gains in some schools, as compared

with control groups, were large; in some they were absent. Overall they were significant. The greatest superiority of the experimental groups was in the discrimination of arguments, as might be expected from our earlier discussion of transfer.

As Watson and Glaser (*1940,* page 32) put it:

> Scores of investigations concur in the conclusion that no subject can be depended upon in itself, as ordinarily taught, to improve the thinking of pupils. On the other hand, almost any subject or project can be so taught as to put pupils on guard against hasty generalization, bias, and the uncritical acceptance of authority.

IMPROVEMENT OF GROUP THINKING

When anyone has a difficult problem to solve, he can often get help from group discussion. Others can contribute promising ideas and can help him evaluate his own ideas. Naturally he will get the most help from people who know something about the problem, but he may also improve his own thinking as a result of talking about the problem to others. Many thinkers have reported that when they are forced to explain the problem to others, they formulate it more adequately for themselves.

If one wishes to improve his own thinking by social facilitation, he should, of course, go to his confreres in a receptive rather than a defensive attitude. Here again the alternation technique can be recommended. He should do the best solitary thinking he can, then seek social help, then think by himself again.

Suppose the group has the problem and the responsibility for settling it. The research summarized in the preceding chapter forces explicit recognition of two different goals of group discussion. If the aim is to keep everyone in the group happy, maximum participation of all members, free communication between them, and a democratic group atmosphere can be recommended. If the aim is to reach the best solution, variability of competence within the group, efficient channeling of communication, and the special role of the expert should be considered.

Once these two aims of group discussion are recognized and the possibility of conflict between them admitted, steps can be taken to bring both together.

If the aim is to maximize the members' satisfaction, they may, if the

problem is clearly formulated, accept the help of an expert in attaining that satisfaction, and even delegate special temporary power to the expert while retaining the right to judge the expert's proposals. After all, an expert is not an expert by fiat; he wins such recognition by a social process.

On the other hand, in many situations in modern society when the group goal is an effective solution, it turns out that this effective solution is one that satisfies the members. This is necessarily the case when the solution must be executed by an investment of the members' time, money, or intellectual support.

In other respects the recommendations for group thinking are the same as for individual thinking: adequate preparation by members and leader, promotion of interest in the goal—in this case the group goal—rather than trivia, progressive practice beginning with easy noncontroversial problems, ventilation of or isolation from intragroup conflicts, and alternation between enthusiastic production of proposals and their critical evaluation.

REFERENCES

Anderson, H. C., Marcham, F. G., and Dunn, S. B., 1944. An experiment in teaching certain skills of critical thinking, *J. Educ. Res.*, *38*:241–251.

Barlow, N. C., 1937. Transfer of training in reasoning, *J. Educ. Psychol.*, *28*:122–128.

Biddle, W. W., 1932. *Propaganda and Education,* Teachers College Contributions to Education, No. 531, New York, Columbia University.

Bloom, B. S., and Broder, L. J., 1950. *Problem-solving processes of college students,* Supplementary Education Monograph, No. 73, Chicago, University of Chicago Press.

Brembeck, W. L., 1949. *The effects of a course in argumentation on critical thinking ability,* Speech Monogr., No. 16, 177–189.

Burack, B., 1950. *The nature and efficacy of methods of attack on reasoning problems,* Psychol. Monog., 64, No. 313.

Gibson, E. J., 1953. Improvement in perceptual judgments as a function of controlled practice or training, *Psychol. Bull.*, *50*:401–431.

Glaser, E. M., 1941. *An experiment in the development of critical thinking,* Teachers College Contributions to Education, No. 843, New York, Columbia University.

Hill, M., 1936. *Training to reason,* Educational Research Series, No. 44, Melbourne, Australia, Melbourne University Press.

Howell, W. S., 1943. *The effects of high school debating on critical thinking,* Speech Monogr. No. 10, 96–103.

Jewett, A., 1940. Detecting and analysing propaganda, *Eng. Jour., 29:* 105–115.

Johnson, D. M., 1953. What the teacher should know about symbols and thought, *Progressive Educ., 30:*145–149.

Kilgore, W. A., 1941. *Identification of ability to apply principles of physics,* Teachers College Contributions to Education, No. 840, New York, Columbia University.

Lefford, A., 1946. The influence of emotional subject matter on logical reasoning, *J. Gen. Psychol., 34:*127–151.

Magaret, A., 1950. Generalization in successful psychotherapy, *J. Cons. Psychol., 14:*64–70.

Melzer, J. H., 1949. Functional logic, *J. Higher Educ., 20:*143–146.

National Society for the Study of Education, 1950. *49th yearbook,* ed. N. B. Henry, Part I, Learning and instruction, Chicago, University of Chicago Press.

National Society for the Study of Education, 1951. *50th yearbook,* ed. N. B. Henry, Part II, The teaching of arithmetic, Chicago, University of Chicago Press.

Osborne, W. W., 1939. Teaching resistance to propaganda, *J. Exper. Educ., 8:*1–17.

Reiner, W. B., 1947. The value of cause and effect analysis in developing ability to recognize cause and effect relationships, *J. Exper. Educ., 15:* 324–330.

Seashore, R. H., and Van Dusen, A. C., 1950. *How to solve your problems,* Chicago, Science Research Associates.

Simpson, B. R., 1923. Training in the technique of thinking as a means of clearer thinking, *School and Society, 18:*358–360.

Szekely, L., 1950. Productive processes in learning and thinking, *Acta Psychologica, 7:*388–407.

Thurstone, T. G., 1949. Learning to think series, *Teacher's manual for the Green Book,* Chicago, Science Research Associates.

Ulmer, G., 1939. Teaching geometry to cultivate reflective thinking: an experimental study with 1239 high school pupils, *J. Exper. Educ., 8:*18–25.

Updegraff, R., and Keister, M. E., 1937. *A study of children's reactions to failure and an experimental attempt to modify them,* Iowa City, University of Iowa Studies in Child Welfare, 13, No. 4.

Watson, G., and Glaser, E. M., 1940. Education for critical thinking, *Advanced School Digest, 6:*29–33.

Wesman, A. G., 1945. *A study of transfer of training from high school subjects to intelligence,* Teachers College Contributions to Education, No. 909, New York, Columbia University.

White, E. E., 1936. A study of the possibility of improving habits of thought in school children by a training in logic, *Brit. J. Educ. Psychol. 6*:267–273.

Winch, W. H., 1923. The transfer of improvement in reasoning, *Brit. J. Psychol., 13*:370–381.

Wrightstone, J. W., 1938. *Appraisal of newer elementary school practices,* Teachers College, New York, Columbia University.

Index of Names

Ach, N., 4, 18, 233, 273
Adams, S., 437, 442
Adamson, R. E., 259
Adkins, D. C., 406, 409, 439, 442
American Institute of Public Opinion, 233
Allport, F., 14, 380, 384
Allport, G. W., 24, 54, 61, 97
Alper, T. G., 376, 389
Anastasi, A., 427, 435, 440
Anderson, G. L., 124, 144, 246, 273
Anderson, H. C., 497, 499
Anderson, H. R., 174
Andrews, E. G., 430, 440
Ansbacher, H., 103, 113
Apicella, F. S., 35, 54
Arai, T., 90, 97
Arnold, L. P., 210, 218
Arnott, P., 149, 177
Asch, S. E., 241–243, 273, 295, 296, 311, 316, 319
Atkinson, J. W., 67, 68, 99, 113, 114

Bach, J. S., 427
Bahrick, H. P., 75, 97
Baker, H. H., 87, 97
Baker, L. E., 295, 319
Baker, L. M., 179, 216
Baker, R., 28, 55, 262, 277
Bales, R. F., 448, 449, 451, 452, 477
Barber, B., 472, 477
Barber, J. L., 362, 384
Barclay, W. D., 187, 216
Barker, R. G., 265, 274, 373, 384, 385
Barlow, J., 87, 98
Barlow, N. C., 484, 497, 499
Barndt, R. J., 382
Barratt, E. S., 395, 440
Bartley, S. H., 89, 97
Battersby, W. S., 257, 273, 274
Becker, M., 340, 383, 388
Bedell, R. C., 140, 143
Beebe-Center, J. G., 303, 307, 319
Bender, M. B., 257, 273, 274
Bendig, A. W., 304, 319
Bensley, M. L., 239, 275
Berenda, R., 296, 298, 300, 319
Berg, E. A., 201, 209, 210, 216, 217
Berg, I. A., 382, 383, 385
Berger, R. M., 425
Berlyne, D. E., 59, 70, 97
Bernyer, G., 412, 440
Berthold, F., 148, 172, 177
Biddle, W. W., 497, 499
Billings, M. L., 139, 143, 392, 394, 440
Binet, A., 2, 3, 17, 103, 113, 409
Bingham, W. V., 311, 319
Birch, H. G., 202, 216, 259, 274, 298, 319
Blake, R. R., 113, 114
Blakey, R. I., 399, 440
Block, H., 295, 311, 319
Bloom, B. S., 30, 54, 483, 499
Blumenfeld, W., 33, 54, 317, 320, 331, 363, 365, 385
Bolles, M. M., 208, 216
Boring, E. G., 1, 20, 306, 320
Botzum, W. A., 409, 440
Bousfield, W. A., 183–188, 191, 192, 206, 216
Bovard, E. W., 463, 477
Bradford, L. P., 472, 477
Bray, C. W., 476, 478
Brembeck, W. L., 497, 499
Brener, R., 92, 97
Brian, C. R., 112, 113
Briggs, A., 96, 97
Broder, L. J., 30, 54, 483, 499
Brotherton, D. A., 247, 274, 333, 358, 385
Brown, D. R., 367, 385
Brownell, W. R., 247–249, 274
Brozek, J., 61, 88, 98, 182, 217
Bruner, J. S., 74, 97, 100, 103, 112–114
Brunswik, E., 302, 320
Bryan, W. L., 4
Bulbrook, M. E., 164, 176

Burack, B., 30, 54, 259, 274, 482, 499
Burt, C., 245, 274, 398, 440
Burtt, H. E., 296, 300, 320
Buss, A. H., 201, 216, 234, 274
Buxton, C., 314, 320

Cady, V. M., 379, 385
Campbell, A. A., 46, 54, 139, 143
Cantril, H., 109, 114, 296, 298, 301, 302, 320
Cardall, A. J., 425, 440
Carrington, D. H., 425, 440
Carroll, J. B., 205, 216, 433, 434, 440
Carter, L., 449, 477
Cartwright, D., 285, 291, 306, 320, 330, 331, 362, 368, 372–376, 382, 385
Cattell, J. M., 17, 285, 321, 369, 386
Cattell, R. B., 24, 54, 96, 97, 205–207, 212, 217, 304, 317, 320, 322, 432, 440, 450, 478
Chant, S. N. F., 291, 320, 373, 385
Chapanis, A., 363, 385
Chapman, D. W., 315, 320
Chein, I., 67, 99
Christensen, P. R., 402, 406, 424, 437, 438, 441, 444
Chute, E., 89, 97
Claperède, E., 28, 54
Clark, R. A., 68, 99, 113, 114
Cofer, C. N., 129–131, 143, 208, 218, 231, 257, 269, 274, 276
Coffin, T. E., 298, 320
Cohen, B. H., 186, 188, 216
Coleridge, S. T., 28
Collier, J. S., 382, 383, 385
Compton, R. K., 92, 97
Comrey, A. L., 402, 441
Cook, T. W., 255, 274
Corter, H. M., 425, 440
Cotzin, M., 378, 386
Coward, A. F., 312, 320
Cowen, E. L., 203, 211–215, 217
Cowley, A., 1
Cox, C. M., 431, 435, 436, 440
Crews, A., 87, 98
Crissy, W. J. E., 318, 320
Cronbach, L. J., 295, 298, 299, 320
Crutchfield, R. S., 71, 100
Culler, A. J., 85, 86, 97
Cureton, E. E., 190, 218
Curran, J. F., 239
Curtis, C. C., 295, 322

Dailey, C. A., 313, 320
Danzig, E. R., 460, 480
Darley, J. G., 476, 478
Darrow, C. W., 77, 97
Darwin, C., 8, 431
Dashiell, J. F., 64, 77, 97, 373, 385, 445, 478
Davidon, R. S., 240, 274
Davis, B. M., 298, 320
Davis, F. B., 400, 402, 419, 420, 440
Davis, R. C., 78, 79, 97, 378, 385
Davis, R. T., 60, 97
Dembo, T., 265, 274
Dennis, W., 268, 278
Descartes, R., 427
Deutsch, M., 456, 478
Deutsche, J. M., 268, 274
Devnich, G. E., 230, 231, 274
Dewey, J., 9, 20, 27, 28, 54, 178, 217, 451
Diethelm, O., 182, 220
Dixon, J. C., 237, 274
Donders, F. C., 5
Doty, R. A., 274
Dougan, C. P., 433, 440
Drever, J. I., 234, 274
Dudek, F. J., 207, 218
Dunbar, D. S., 446, 479
Duncker, K., 160, 161, 176, 194, 213, 217, 250–252, 258, 259, 274
Dunn, S. B., 497, 499
Dunning, G. M., 418, 440
Durkin, H. E., 31–33, 54, 255

Ebbinghaus, H., 11
Ehrmann, J. C., 299, 300, 322
Eindhoven, J. E., 262, 274
Einstein, A., 431
Eisenstein, D., 214, 217
Elliott, D. N., 179, 216
Elliott, M. H., 163, 176
Ellis, H., 436
Ellis, W. D., 113
English, H. B., 46, 54, 137, 143, 235, 275
Ericksen, C. W., 214, 217
Ericksen, S. C., 256, 275
Euler, L., 245, 260
Everall, E. E., 202, 217
Ewert, P. H., 241, 275

Eysenck, H. J., 303, 317, 320, 469, 478
Eysenck, M. D., 303, 321

Falkenburg, D. R., 296, 320
Farmer, E., 379, 385
Farnsworth, P. R., 290, 295, 321, 323
Faust, W. L., 466, 479
Fechner, G. T., 15
Feder, D. D., 419, 441
Ferguson, G. A., 212, 219
Ferguson, L. W., 290, 321
Fernberger, S. W., 362, 385
Festinger, L., 285, 306, 315, 320, 321, 368, 370–373, 376, 385, 386
Fey, E. T., 210, 217
Findley, W. G., 174
Fiske, D. W., 379, 387, 414, 442
Fitts, P. M., 75, 97
Flanagan, J. C., 428, 441
Flurry, C., 239, 279
Foley, J. P., 129–131, 143, 172, 176, 427, 435, 440
Forlano, G., 290, 323
Forsyth, E., 174
Fouriezos, N. T., 456, 478
Fox, J., 214, 218
Frawley, P. J., 383, 386
Freeman, G. L., 78, 79, 87, 97, 378, 386
French, J. R. P., 472, 477
French, J. W., 96, 98
Freud, S., 12, 13, 427, 434
Frick, F., 296, 321
Frost, R., 280
Fruchter, B., 205, 217
Fryer, D., 91, 98
Fullerton, G. S., 285, 321, 369, 386
Furneaux, W. D., 317, 320

Gallenbeck, C., 376, 386
Galton, F., 17
Gans, R., 419, 441
Garner, W. R., 362, 384
Garrett, H. E., 137, 143, 371, 379, 386
Gengerelli, J. A., 240, 275
Ghiselin, B., 167, 176, 261, 266, 275
Gibb, C. A., 205, 217
Gibbons, H., 235, 275
Gibby, R. G., 69, 98
Gibson, E. J., 495, 499
Gibson, J. J., 64, 74, 77, 98
Giese, W. J., 378, 386
Glaser, E. M., 407, 441, 444, 485, 497, 498–500
Goethe, J. W., 427, 431
Goldstein, K., 18
Goodenough, F. L., 112, 113, 128, 143, 173, 176
Goodman, C. C., 103, 113
Goodman, C. H., 410, 412, 444
Goodman, L., 430
Goodstein, L. D., 214, 215, 217
Gordon, K., 467, 478
Gottschaldt, K., 7, 104, 105, 113
Gould, R., 315, 321
Gramlich, F. W., 149, 177
Grant, D. A., 201, 204, 209, 217, 219, 239, 275
Green, R. F., 402, 406, 441
Greenberg, A., 200, 218
Greene, E. B., 379, 384, 386
Gritten, F., 384, 386
Guetzkow, H., 35, 54, 61, 88, 98, 182, 207, 210, 211, 213, 217, 447, 448, 456, 472, 478, 479
Guilford, J. P., 94, 98, 111, 137, 144, 271, 275, 290, 318, 321, 354, 378, 386, 400, 402, 406, 408, 409, 412, 413, 420–424, 432, 437, 438, 441, 444
Gurnee, H., 470, 471, 478

Haight, D., 307, 323
Hall, C. S., 137, 144
Hall, K. R. L., 70, 71, 98
Hamilton, G. V., 39, 202, 217
Hamilton, J. A., 202, 217
Hanfman, E., 18, 208–210, 217
Hargreaves, H. L., 204, 217, 432, 441
Harlow, H. F., 38, 54, 59, 60, 97, 98, 201, 218
Harlow, M. K., 59, 60, 98
Harman, H. H., 401, 442
Harmon, F. L., 87, 98
Harrell, T. W., 271, 275, 395, 442
Harrison, E. S., 299, 313, 321
Hartman, D. A., 380, 384
Hartmann, G. W., 301, 321
Hartocollis, P., 182, 183, 218
Haythorn, W., 477
Heidbreder, E., 234, 238–240, 275, 404
Heineman, C., 179, 219
Heintz, R. K., 365, 386
Heise, G. A., 454, 478

Helmholtz, H., 9, 27, 167, 477
Helson, H., 342, 347, 348, 353, 359, 360, 386, 387
Hemphill, J. K., 450, 478
Henderson, M. T., 87, 98
Henmon, V. A. C., 372, 386
Herrick, C. J., 152, 167
Hertzka, A. F., 406, 424, 441
Hertzman, M., 295, 311, 315, 319, 321, 379, 386
Hess, J., 211–213, 216, 217
Heyns, R. W., 35, 54, 447, 456, 472, 478, 479
Hilgard, E., 118
Hill, M., 484, 499
Hinckley, E. D., 290, 321, 363, 386
Hines, R., 239, 275
Hobhouse, L. T., 10
Holley, J. W., 318, 321
Hollingworth, H. L., 9, 16, 230, 275, 276, 285, 321, 353, 386
Holzinger, K. J., 401, 442
Hovey, H. B., 87, 98
Hovland, C. I., 235, 239, 276, 290, 295, 321
Howell, W. S., 497, 499
Hughes, R. H., 180, 189, 220
Hull, C. L., 121, 144, 187, 218, 233, 234, 240, 246, 276, 434
Humphrey, G., 6, 20
Hunt, J. McV., 208, 218, 269, 276
Hunt, W. A., 341, 367, 387
Husband, R. W., 253, 276, 465, 478
Huston, P. E., 92, 98
Hutchinson, E. D., 167, 176, 261, 263, 265, 276
Hutt, M. L., 456, 478
Huxtable, Z. L., 90, 98

Immergluck, L., 72, 99
Ipsen, G., 354, 387
Irion, A. L., 137, 144
Israel, H. E., 71, 100, 211, 219, 231, 277
Ivy, M., 239, 275

Jacobs, R., 411, 442
James, W., 8, 9, 51
Janis, I. L., 296, 321
Janis, M. G., 131, 143
Jenkins, J. J., 188, 218
Jenness, A., 462, 469, 478
Jensen, E. M., 39, 54
Jersild, A. T., 86, 99, 151, 177
Jewett, A., 497, 500
Johanson, A. M., 77, 99
Johnson, D. M., 96, 97, 142, 144, 185, 205, 206, 218, 225, 247, 276, 281, 285, 290, 299, 300, 306, 310, 322, 331, 332, 337, 342, 343, 346, 348–353, 369–373, 376, 378–380, 384, 386, 387, 417, 420, 432, 442, 472, 478, 487, 500
Johnson, R., 430
Johnson, R. C., 185, 206, 218
Jones, L. V., 411, 442

Kaplan, E., 235, 279
Karsten, A., 89, 99
Karwoski, T. F., 148, 149, 172, 177
Kasanin, J., 18, 208–210, 217
Katona, G., 124, 144
Katzell, R. A., 79, 99
Kaufman, E. L., 38, 39, 46, 54, 338, 340, 363, 364, 387, 388
Keister, M. E., 485, 491, 500
Kellogg, W. N., 373, 381, 387
Kelly, E. L., 379, 387, 414, 442
Kelman, H. C., 295, 322
Kendall, W. E., 79, 97
Kendler, H. H., 200, 218
Kennedy, J. L., 88, 99, 100
Kepler, M. O. 185, 219
Kerr, W. A., 437, 444
Kettering, C. F., 221
Kettner, N. W., 406, 441
Kilgore, W. A., 486, 500
Killian, C. D., 46, 54, 137, 143
King, G. F., 299, 300, 322
Kirkpatrick, J. J., 190, 218
Kleemeier, L. B., 87, 99
Kleemeier, R. W., 88, 99, 207, 218
Klein, G. S., 384, 387
Koffka, K., 7, 103, 110, 114
Köhler, W., 5, 7, 10, 31
Koob, H. F., 245, 278
Krechevsky, I., 202, 217
Külpe, O., 2–7, 20, 74, 99
Kuo, Z. Y., 234, 276

Lambert, J. F., 241, 275
Lambert, W. E., 466, 479
Landahl, H. D., 306, 322

Landis, C., 208, 216
Lange-Eichbaum, W., 435, 442
Langsam, R. S., 419, 442
Lanier, L. H., 476, 478
Lantz, B., 182, 218
Lanzetta, J., 477
Leavitt, H. J., 453–455, 479
Le Bon, G., 13
Leeds, D. S., 200, 219, 220
Leeper, R. W., 26, 54
Lefford, A., 297, 322, 491, 500
Levine, A. S., 414, 442
Levine, R., 67, 99
Levitt, E. E., 215, 218
Lewin, K., 7, 265, 274, 472, 479
Lewis, D. J., 437, 441
Lewis, H. B., 298, 315, 316, 321, 322
Lincoln, A., 492
Lindsley, D. B., 77, 99
Linton, R., 21
Long, L., 137, 145, 182, 220, 236, 237, 240, 276–279, 439
Lord, M. W., 54
Lorge, I., 295, 322
Lowes, J. L., 28, 54
Luborsky, L. B., 304, 317, 320, 322
Luchins, A. S., 198–201, 211–214, 218, 231, 258, 272, 276
Luchins, E. H., 200, 218
Lund, F. H., 296, 322
Lurie, W., 295, 322
Lyerly, S. B., 406, 409, 440

Macaulay, T. B., 431
McCarthy, D., 269, 276
McCartor, M. A., 90, 98
McClelland, D. C., 35, 54, 67, 68, 99, 113, 114
McCurdy, H. G., 466, 479
McDougall, W., 12
McGarvey, H. R., 291, 322, 366, 367, 387
McGeoch, J. A., 118, 120, 137, 144
McGinnies, E., 67, 99, 112, 114
McGregor, D. M., 296, 298, 322
Mackworth, N. H., 88, 99
MacMillan, Z. L., 172, 176
Madden, E. H., 256, 278
Magaret, A., 490, 500
Maier, N. R. F., 8, 39–41, 46, 55, 72, 99, 119, 121–124, 127, 138, 139, 144, 202, 213, 218, 256–259, 273, 276, 396, 458, 459, 479
Maltzman, I., 200, 214, 218
Mandell, M. M., 425, 432, 437, 442
Mandell, W., 295, 321
Marcham, F. G., 497, 499
Mark, A. L., 185, 206, 218
Marks, M. R., 164, 165, 177, 234, 250, 276
Marple, C. H., 296, 300, 320
Marquart, D. I., 202, 203, 210, 211, 218
Marquis, D. G., 447, 456, 472, 479
Mathews, M. A., 131, 143
Matin, L., 409, 442
May, M. A., 4
Meili, R., 8, 20, 408, 442
Meirowitz, B., 477
Melton, A. W., 118, 120, 144
Meltzer, B. N., 437, 442
Meltzer, H., 268, 277
Meltzoff, J., 69, 99
Melzer, J. H., 497, 500
Meyer, D. R., 59, 60, 98
Michels, W. C., 360, 387
Miles, C. C., 128, 145, 173, 177
Miller, G. A., 454, 478
Miller, J. G., 295, 322
Miller, W. S., 414, 442
Milton, J., 29
Mintz, A., 456, 457, 479
Molteno, E. V., 206, 217
Montague, W. P., 56
Moore, H. T., 14, 296, 300, 322
Moore, J. H., 74, 77, 99
Morgan, C. L., 10
Morgan, J. J. B., 87, 99
Morrisett, L., 200, 214, 218
Morse, H. T., 174
Mosier, D. I., 291, 322
Mosteller, C. F., 301
Mt. Holyoke Psychophysical Research Unit, 362, 387
Mowrer, O. H., 77, 78, 99
Muller, G. E., 15
Murphy, G., 1, 14, 20, 67, 99, 103, 114, 172, 177, 445, 464, 476, 479
Murphy, L. B., 445, 465, 479
Myers, C. T., 439, 442

Nash, M. C., 360, 387
National Society for the Study of Education, 174, 177, 487, 500
Newcomb, T. M., 445, 465, 479
Newton, I., 115

Nissenson, M., 230, 231, 277
Norris, E. D., 373, 388

Öbrink, J., 379, 384, 388
Odbert, H. S., 24, 54
Oden, M., 393, 431, 443
Ogburn, W. F., 434, 443
Ogden, R. M., 20
Olander, H. T., 246, 277
Oliver, J. A., 212, 219
Ordan, H., 268, 277
Osborne, W. W., 497, 500
Oseas, L., 241, 277
Osgood, C. E., 302, 322
Overman, J. R., 246, 277
Overstreet, P., 239, 275

Page, R., 108, 109, 114, 318
Painter, W. S., 90, 99
Pascal, B., 431
Paschal, F. C., 64, 73, 74, 99
Paterson, D., 414, 444
Patrick, C., 28–31, 55, 168, 169, 177, 178, 219, 261–264, 277
Patrick, J. R., 39, 40, 55, 202, 219
Pavlov, I. P., 11
Pemberton, C., 409, 443
Philip, B. R., 333–337, 339, 347, 354, 356, 360, 361, 363, 388
Piaget, J., 142, 268, 269, 277
Pillsbury, W. B., 81, 100
Pintner, R., 290, 323
Platt, W., 28, 55, 262, 277
Poffenberger, A. T., 89, 100
Poincaré, H., 9, 167
Postman, L., 71, 74, 97, 100, 108, 109, 112, 114, 318, 347, 388
Poulton, E. C., 77, 100
Pratt, K. C., 247, 274, 333, 358, 385
Preston, M. G., 467, 479
Proshansky, H., 103, 114

Rabinowitz, H. S., 259, 274
Radin, P., 13, 269, 277
Ramond, C. K., 234, 250, 276
Rankin, R. E., 75, 97
Rapaport, D., 208, 209, 219
Raskin, E., 436, 443
Razran, G., 129, 144, 152, 177, 181
Read, J. M., 247, 274, 333, 358, 385
Reed, H. B., 37, 55, 234, 239, 241, 277
Reed, J. B., 371, 372, 388
Rees, H. J., 71, 100, 211, 219, 231, 277
Reese, E. P., 54
Reese, T. W., 54, 338, 340, 363, 364, 387, 388
Regan, J. J., 318, 320
Regensburg, J., 307, 323
Reid, J. W., 162, 177
Rein, W. C., 274
Reiner, W. B., 497, 500
Reitz, W., 33, 55
Rethlingshafer, D., 96, 100, 363, 386
Reynolds, F., 205, 218
Rhoades, C., 379, 387
Richardson, H. M., 253, 254, 277
Richman, H., 200, 218
Riess, B. F., 129, 144
Riggs, L. A., 92, 98
Rimoldi, H. J., 408, 443
Roberts, M. L., 373, 388
Robinson, E. E., 89, 100, 198, 219
Rodnick, E. H., 92, 100
Roe, A., 436, 437, 443
Roff, M., 459–461, 480
Rogers, S., 330, 338, 340, 363, 366, 387, 388
Rokeach, M., 201, 211, 214, 215, 219, 270, 277
Rosen, G. P., 208, 216
Roslow, S., 138, 144
Ross, B. N., 204, 219
Rossman, J., 28, 55, 262
Rotter, J. B., 201, 219
Rowell, N. M., 131, 143
Ruger, H., 9, 160, 177, 261, 277
Rundquist, E. A., 383, 388
Rupel, J. W., 204, 219
Russell, R. W., 268, 278
Russell, W. A., 188, 218
Ryans, D. G., 95, 96, 100

Saadi, M., 295, 323
Saffir, M. A., 290, 323
Salter, M. D., 291, 320, 373, 385
Sanford, R. N., 67, 100
Sanua, V., 418, 443
Sargent, S. S., 229, 230, 231, 277, 278
Sartain, A. Q., 425, 443
Scates, D. E., 174
Schachter, J., 149, 177

Schiff, E., 433, 440
Schlosberg, H., 179, 219
Schneck, M. R., 137, 143
Schmierla, T. C., 119, 144
Schoenfeld, N., 384, 387
Schopenhauer, A., 390
Schroder, H. M., 201, 219
Seashore, R. H., 489, 500
Sedgewick, C. H. W., 183–188, 191, 192, 206, 216
Seeman, W., 383, 388
Sells, S. B., 245, 278, 279, 294, 299, 310, 323
Selz, O., 5, 6
Settlage, P. H., 60, 97
Seward, G., 369, 378, 388
Seward, J. P., 121, 144
Shafer, R., 103, 114
Shakespeare, W., 427
Shakow, D., 92, 98, 100
Shannon, J. R., 435, 443
Shartle, C. L., 142, 144
Shaw, M. E., 465, 466, 479
Sherif, M., 109, 110, 114, 290, 296, 321, 323, 463, 479
Shipley, W. C., 373, 388
Siipola, E. M., 69, 100
Simpson, B. R., 484, 497, 500
Smith, B. B., 468, 479
Smith, E., 415, 443
Smith, H. C., 417, 420, 442, 446, 472, 478, 479
Smith, K. U., 376, 386
Smoke, K. L., 233–235, 278
Solomon, M. D., 214, 219
South, E. B., 457, 479
Spearman, C., 17, 398, 402, 405, 443
Spence, K. W., 234, 278
Stacey, C. L., 124, 144
Stagner, R., 302, 322
Stead, W. H., 142, 144
Stretch, L. B., 249, 274
Strodtbeck, F. L., 452, 477
Stroop, J. R., 467, 479
Studman, L. G., 182, 205, 219
Supa, M., 81, 86, 101
Swenson, E. J., 124, 144, 246, 278
Swineford, F., 401, 442
Sylvester, A., 81, 100
Symonds, P. M., 33, 55
Szekely, L., 124, 145, 486, 487, 500

Tarde, G., 13
Taubman, R. E., 340, 388
Taves, E. H., 38, 55
Taylor, C. W., 205, 219
Taylor, D. W., 466, 479
Terman, L. M., 128, 144, 173, 177, 204, 219, 393, 431, 443
Teuber, H. L., 257, 273, 274
Thiele, C. L., 246, 278
Thomas, W. F., 92, 100
Thompson, G. G., 185, 214, 217, 219
Thorndike, E. L., 10, 31, 90, 304, 310, 323, 419, 443
Thorndike, R. L., 471, 479
Thornton, G. R., 96, 100, 296, 299, 307, 308, 323
Thurstone, L. L., 17, 37, 55, 111, 114, 137, 145, 205, 219, 266, 278, 398–401, 405, 408, 410–412, 420, 443
Thurstone, T. G., 111, 114, 137, 145, 205, 219, 399, 410–412, 443, 485, 500
Tiner, L. G., 212, 217
Tinklepaugh, O. L., 163, 177
Titchener, E. B., 2, 3, 20
Tolman, E. C., 8, 404
Toops, H. A., 309, 323
Travis, R. C., 88, 99, 100
Tredick, V. D., 410
Tresselt, M. E., 113, 114, 200, 219, 220, 340, 349, 352, 383, 388
Trow, W. C., 384, 388
Turchoie, R. M., 354, 388
Tyler, R. W., 415, 443
Tylor, E. B., 13

Ulmer, G., 485, 500
Underwood, B. J., 120, 145, 180, 189, 220, 241, 252, 277, 288
Updegraff, R., 485, 491, 500

Vanderplas, J. M., 113, 114
Van Dusen, A. C., 489, 500
Van Zelst, R. H., 437, 444
Vernon, M. D., 408, 412, 444
Vernon, P. E., 48, 55, 206, 220, 266, 278, 405, 444
Vigotsky, L. S., 18, 208, 220
Vinacke, W. E., 20, 142, 145, 233, 262, 268, 274, 278
Volkmann, J., 54, 315, 320, 338, 340, 341, 362, 364, 367, 378, 387, 388

Wallas, G., 9, 27–30, 51, 55, 168, 177, 178, 220, 261, 278
Wapner, S., 376, 389
Warden, C. J., 59, 101, 253, 278
Washburn, M. F., 307, 323
Watson, G. B., 407, 444, 464, 465, 480, 498, 500
Watson, J. B., 10, 11
Watt, H. J., 4
Watters, G. V., 414, 444
Weaver, H. E., 256, 278
Wegrocki, H. J., 269, 278
Weigl, E., 18, 208, 220
Weislogel, M. H., 430, 432, 444
Weiss, W., 235, 276
Welborn, E. L., 46, 54, 137, 143
Welch, L., 120, 137, 145, 182, 220, 234, 236, 237, 240, 276, 278, 279, 433, 439, 440
Weld, H. P., 459–461, 480
Welford, A. T., 176, 177
Wells, F. L., 150, 152, 175–177
Wenzel, B. M., 239, 279
Werner, H., 7, 235, 268, 271, 272, 279
Wertheimer, M., 6, 7, 70, 101, 103, 104, 114, 161, 177, 213, 251, 252, 254, 279
Wesley, E. I., 215, 220
Wesman, A. G., 484, 500
Wever, E. G., 331, 389
White, E. E., 484, 497, 501
White, M. H., 90, 98
White, R. K., 436, 444
Whitfield, J. W., 256, 279
Wiener, M., 211–213, 215, 217
Wilkins, M. C., 244, 250, 279
Williams, A. C., 295, 323
Williams, F., 301
Wilson, G. M., 250, 279
Wilson, J. T., 376, 478
Wilson, R. C., 437, 438, 441, 444
Winch, W. H., 485, 501
Winder, A. E., 212, 217
Wispe, L. G., 450, 478
Witryol, S. L., 299, 323
Wittenborn, J. R., 93, 94, 101
Wolfle, D., 266, 279
Woodrow, H., 76, 78, 79, 101, 354, 389
Woodworth, R. S., 2, 3, 5, 12, 20, 65, 92, 101, 104, 114, 148, 153, 167, 177, 179, 180, 220, 222, 245, 252, 261, 279, 294, 323, 324, 330, 332, 352, 362, 389
Wright, M. E., 457, 465, 480
Wrightstone, J. W., 497, 501
Wundt, W., 2, 3, 6
Würzburg School, 3, 5, 6, 9, 12, 18

Yerkes, R. M., 10
Young, C. W., 81, 86
Young, P. T., 92, 97, 100, 101
Youtz, R. P., 200, 220

Zelen, S. L., 215, 218
Zener, K. E., 331, 389

Index of Subjects

Ability, 270, 271, 390–444
  perceptual, 111
  verbal, 267
Abstract, attitude, 18
  judgment, 288–291, 296, 298
  scale, 246
Abstraction, and difficulty of concepts, 239, 240
  and flexibility, 272, 273
  attitude toward, 485
    See also Cognitive patterns; Concepts; Conceptualization
Adaptation level, 342, 347, 348, 353, 359–361
Additivity assumption, 306–308
Affective judgment, 288, 296, 303, 304, 341
  preference, 135, 136
Alcohol, and creative thinking, 266
  and fluency, 182, 183
Alertness, 88
Alpha rhythm, 77
Alternation technique for solving problems, 265, 488, 492, 498
Anagram problem, 229–231
Anchoring, 362–367, 383
Anxiety, and extremism, 383
  and fluency, 182
  and rigidity, 214, 215
Apprehension span, 80, 92
Arithmetic problem solving, 247–252, 394, 401
  improvement of, 485
Association experiment, 147–155, 178
Associationism, 2, 5, 9
Atmosphere effect, 294, 299
Attention, 92–95
  and set, 73–75
  span, 80–83, 92
    See also Set
Attitude, and confidence, 369, 380, 381
  and set, 66
  of thoughtfulness, 485
  scales, 290, 291
Attributes, 135
Aufgabe, 4, 5
Authoritarianism, and rigidity, 215

Behaviorism, 10–12
Bewusstseinslagen, 3
Biological motives, 57

Category, boundary, 330–337
  limen, 330–337
  of judgment, 326–330
  threshold, 330–337
  threshold and confidence, 370, 371
  threshold and judgment time, 373–376
Central tendency of judgment, 16, 353
  as regression, 354–359
Centrality of group structure, 454, 455
Certainty, *see* Confidence
Check list, 309
Closure, 408, 409
Clustering of words, 187
Cognition and perception, 102
Cognitive patterns, 6–8, 222–226, 232–245, 253, 270
  comparison of, 312, 313
Communication, and judgment, 283, 324
  theory, 239, 256
  within groups, 445–455
Comparative psychology, 9
Concepts, 133, 134, 142
  formation of, 231–241
  systems of, 236, 237
Conceptualization, 267–270
Confidence of judgment, 368–371, 383, 384
  and accuracy, 379, 384
  and atypical opinion, 379–381
  and time, 378, 379
  correction for, 384
Conscious thought, 75
Consciousness, 2, 6, 19, 25

Context, learning words from, 235
of judgment, 314–317, 329, 367
Continuous association, 181–188
Control of emotion, 490
Correlational analysis, 44–48, 266, 302–304
of groups, 450
Correspondence principle, 341, 342
Creative thinking, 29, 30, 260–266, 427–439
characterized, 166, 167
preparation for, 165–171
Critical incidents, 428–430
Critical thinking, 407, 408, 415, 497, 498
Curiosity, 59, 156

Decision, *see* Judgment
Deduction, 243–245, 397–399, 405, 409, 412
Denkpsychologie, 6, 8, 15
Determining tendency, 4, 151
See also Set
Difficulty, of abstract judgments, 289
of arithmetic problems, 249–252
of concept formation, 239–241
of judgment, 376–379
of problems, 190, 229, 260
of spatial and mechanical problems, 253–260
Dimension, 135
as figure, 304–306
Dimensional analysis, 120
Direction, 72, 458
Discrimination, 146, 147
methods, 227, 228, 232, 233
Distraction, 86, 87, 152
and group thinking, 457
Dogmatism and rigidity, 215
Dynamics of thought, 36–101, 455–459
improvement of, 485

Education and reasoning, 412–421
Eduction, *see* Reasoning
Egocentrism, 269
Ego, involvement, 169–171
motives, 57, 58, 62, 73; and creativity, 437
Einstellung, 3, 4, 12, 13, 151, 199
Electroencephalograph, 77, 79
Embedding, 104, 105, 240, 258, 265
Eminence and creative accomplishment, 427, 435, 436
Emotion, and constructiveness, 265, 266
and flexibility, 202–204
and personal problems, 170
Equivalent stimuli, 122, 123
Estimating, 39, 46
ability, 421, 423, 424
and judging, 284
Ethnocentrism, 214, 270
Evaluation, 264
abilities, 424, 425
by groups, 452, 477
Extremism of judgment, 382, 383
and confidence, 380, 381

Factor analysis, 48–50, 303, 394–396, 399–407, 419, 420
integration of, 406, 407
Familiarity, and creative thinking, 265
and difficulty of problems, 244, 249
Fantasy, 57, 74
Fatigue, 88–90, 116, 117
Figure and ground in judgment, 304–306
Fixation, 197
See also Flexibility
Flexibility, 95, 197–204, 207–216, 437, 438, 492
of closure, 408, 409
of judgment, 496
Fluency, 181–188, 204–207, 432, 433, 437, 438
Forethoughtfulness, 176
Formulation, 29, 30
improvement of, 487–489
Frame of reference, 7, 16, 102, 108–110, 314–317
contrasted with set, 109, 110
Frustration, and constructiveness, 265, 266
and flexibility, 202–204
Function, 8, 9
of judgment, 281–285
Functionalism, 8

General impression, 309–312
General intelligence, 111, 137–140, 142, 214, 266, 267, 270, 393–396, 400–402
and creativity, 431, 432
and reasoning, 409–412, 418
Generalization, 134, 135, 180, 181
ability, *see* Concepts; Conceptualization
of conditioning, 129, 152

of number relationships, 246
theory, 342–347, 352
Gesamtaufgabe, 5
Gestalt psychology, 6–8, 16, 103, 161, 409
Good judgment, 299, 305, 306, 318, 319, 421–427
Group homogeneity, 461–463, 469–471
Group thinking, 445–480
improvement of, 498, 499

Halo effect, 281, 310, 311, 318, 495
Hedonic judgments, 288, 296, 303, 304
Hedonic summation, 307
Homogeneity of judgment, 461–463, 469–471
Hypotheses, 234, 238
and reasoning, 406

Ideas, 22, 133
Ideational fluency, 133, 205, 433, 437, 438
Ideational learning, 133
Illumination, 9, 28–30, 102, 178, 263, 264
Imageless thought, 3
Immediate memory, and thinking, 82, 83, 92
span, 80–83, 92
Improvement of thinking, 481–501
Incubation, 9, 13, 27, 28, 30, 261, 262, 266
Individual differences, 43–50, 91–97, 111–113, 390–444
in dynamics, 91–97
in flexibility, 207–216
in fluency, 204–207
in judgment, 317–319, 382–384
in learning, 136–143
in manipulating patterns, 266–273
in perception, 111–113
in preparation, 171–176
Induction, 397–399, 403, 405, 408, 409, 412
See also Concepts; Deduction; Reasoning
Information, 132
and good judgment, 426
and problem solving, 139–142
theory, 239, 256
Insight, 10, 36, 37, 102, 195, 196, 229, 230
Instrumental patterns, 223–226, 238, 247, 253
Integration tests, 94, 95
Intelligence, *see* General intelligence
Intrinsic motivation, 58–62, 156, 223, 233
Introspection, 2, 25, 28
systematic experimental, 4
Intuitive judgments, 295, 304

Judgment, 19, 51, 116, 117, 280–388
as an ability, 400, 421–427
by social groups, 459–463, 467–472
improvement of, 494–497
Jury decisions, 459–461

Leadership and group thinking, 458, 459, 466
Learning, 115–145
and creativity, 433, 434
and later problem solving, 486, 487
to think, 485
transfer of, 484–486
Levels of analysis, 52
Logic of measurement contrasted with psychology of judgment, 330
Logical operations, 397, 398, 405
See also Reasoning; Syllogism
Logical problems, 228–245, 396–399

Manipulation drive, 60
Marbe's law, 149, 179, 186
Maturation, 116, 117
Mechanical problems, 252–259
Mediated generalization, 129–131, 180
Memory factor, 137–139
Motivation, 12, 56–63, 91, 156
and concept formation, 233, 234
and perception, 103, 104
and set, 68, 69, 163
See also Dynamics
Muscular tension, 78, 79, 87, 88
and creative thinking, 263

Negative instances in concept formation, 235
Numerical ability, 270, 271, 394
Numerical problems, 245–252
Numerousness, 38, 46

Objective experimentation, 36–43, 46
Objectivity of judgment, 318
Oddity method, 232
Opinion polls, 301, 302
Originality, 437, 438

Patterns, and problem solving, 228–260
closure of, 408, 409
description of, 226–228
of communication, 453–455
of personality, 241–243
organization of, 222–226
Penetration, 160
Perception, 102–114, 408, 409
and judgment, 284, 285
and set, 73–75
Perceptual judgments, 287, 302
Perceptual organization, 126, 127
Perceptual patterns, 222, 223
Perseveration, 95, 202, 204, 207, 211
Persistence, 95–97
and creative thinking, 266, 436
and personal problems, 170
Personal problems, 169–171, 250
Personality, and creativity, 435–437
and judgment, 383
and judgment of jokes, 317, 318
and rigidity, 214, 215
pattern of, 241–243
Play, 59, 204
Point of subjective equality, 16, 331
Practice, and complexity, 256
and concepts and principles, 239–241
and creative thinking, 265
and flexibility, 198–201
and formulation of problems, 488
and imagination, 493
and judgment, 495, 496
Preparation, 4, 9, 27, 29, 30, 51, 76, 146–177, 260, 261
for judgment, 282, 285–287, 292, 293, 325, 326
for social problem solving, 474–476
improvement of, 486–491
time, 175, 176, 225, 226
Preutilization, 258, 259
Primitive thought, 13, 269
Principles, 237, 238, 486, 487
Problem, definition of, 63
Problem solving, 19
abilities, 205, 392–396
analysis of, 228–260
and judgment, 283, 284
by groups, 445–480
contrasted with learning, 121–126, 138–140
definition of, 63
preparation for, 154–165, 173–176, 225, 226
tests, 139, 140
Production process, 51
and judgment, 317
complex patterns and problems, 221–279
in groups, 463–467, 476
improvement of, 491–494
simple responses, 178–220
Productive thinking, 122, 178
Protocols, analysis of, 34, 35
Psychoanalysis, 13
Psychopathology, 13
and creativity, 435, 436
Psychophysics, 15, 75, 280, 283
Psychophysiology of set, 77–80

Range of equivalence, 331, 375, 376
Rating method, 339
Rational learning, 137, 138
Rationalization, 57
Reaction time, 75–78, 88, 91, 92
Reading and reasoning, 419–421
Reasoning, 243–245, 313, 396–421
and judgment, 425
and learning, 117, 121, 122, 138–140
Recognition and judgment, 285
Reference level, 107
Reference point, 106
Reference scale, 107, 108, 113
Regression theory of central tendency, 354–359
Reproduction methods, 227, 232, 233, 337
Research administration, 475, 476
Response patterns, 222, 226, 227
Response preferences, 361, 362
Response scale, 328–367
Response sets, 295, 299
Revision, 29, 30, 264
Rigidity, 197, 212–215
See also Flexibility

Satiation, 88–90
Scale, of abstract attributes, 135, 290, 291
of judgment, 328–367
of reference, 107, 108, 113
Scheduling of problems, 491
Schema, 5, 153, 154
See also Search model
Schizophrenic thought, 210, 269

Search, 225, 226
  model, 153–155, 188–191, 225, 226, 238, 397
Semantic conditioning, 129, 224
Semantic generalization, 129–132, 152, 180, 181
Sensitivity and creativity, 434, 436–438
Serial judgment, 308, 309
Set, 64–95, 147–176
  and conceptualization, 268
  and production of words, 188–191
  contrasted with abstraction, 208, 209, 272, 273
  contrasted with frame of reference, 109, 110
  definition of, 65
  for continuous intellectual work, 83–90
  for discovering concepts, 241
  for judgment, 283, 292–294, 325, 371
  for single stimulus and response, 64–80
  for stimulus dimension, 304–306
  mental, 4
  multiple, 74
  perceptual, 111–113
  recovery from, 200, 211, 213, 215
  response, 295
  shift of, 191–204, 207–216, 258
  susceptibility to, 200, 211, 213, 215
Sex differences, 128, 140, 383
Social interaction and set, 68, 163–165
Social motives, 57, 58
Social organization of thought, 472–477
Sorting tests, 208–210
Spatial abilities, 271, 395
Spatial problems, 252–259, 395
Spectator behavior, 238
Stereotypy, 197, 202, 203, 210
  See also Flexibility
Stimulus control, 288, 290, 291
Subjective procedures, 25–36, 39–41, 249
Subtizing, 39, 46
Subtraction method, 5
Successive hurdles, 309
Suggestion, 68, 69, 295, 316, 317
Surgency, 206
Syllogism, 3, 138, 244, 245, 294, 296–298, 310, 313, 398, 405, 409

Task-induced set, 72, 153, 165, 195, 196, 198
Thinking aloud, 28
Time of judgment, 371–376, 384
  and accuracy, 379
  and confidence, 378, 379, 384
Time-order error, 347
Tools, 223, 224
Transfer of improvement, 484–486
Trial and error, 10, 33, 36, 37, 195, 196, 229, 230, 396
Trifunctional analysis, 50, 51, 473, 474

Verbal ability, 267, 270, 395, 410, 420, 426
  training in, 497
Verbal aids in problem solving, 253, 257, 260, 396
Verbal problems, 228–245
Verification, 9, 28, 29, 234
Vicarious trial and error, 404
Vigilance, 75, 88
  See also Attention; Set
Visualization ability, 271
Vocabulary, 128, 133, 142, 417, 418

Word association, 147–155, 171–173, 179–181, 438
Words, 127–132
Work planning, 423, 424
Work samples, 430

Zeitgeist, 475