Part II

Foundations of Beliefs, Attitudes, Intentions, and Behaviors

Chapter 5

Belief Formation

The centrality of the belief concept has been emphasized in previous chapters. Within our conceptual framework, beliefs about an object provide the basis for the formation of attitude toward the object, and we have shown that attitudes are usually measured by assessing a person's beliefs. Although the importance of beliefs has frequently been acknowledged (Thurstone, 1931; Cronkhite, 1969; Ostrom, 1968; Rokeach, 1968; Scheibe, 1970), surprisingly little research in the attitude area has focused on the acquisition or formation of beliefs. Clearly, in order to account for the formation and change of attitudes and intentions, the processes of belief formation have to be investigated. Generally speaking, beliefs refer to a person's subjective probability judgments concerning some discriminable aspect of his world; they deal with the person's understanding of himself and his environment. Specifically, we have defined belief as the subjective probability of a relation between the object of the belief and some other object, value, concept, or attribute.[1] Thus a person may believe that he possesses certain attributes (e.g., that he is intelligent, honest, punctual, etc.), that a given behavior will lead to certain consequences, that certain events occur contiguously, etc.

This definition implies that belief formation involves the establishment of a link between any two aspects of an individual's world. One obvious source of

1. A distinction can be made between beliefs in the existence of an object (e.g., belief *in* God) and beliefs in the existence of a relationship linking the object to some attribute (e.g., a belief *about* God, such as "God is omnipotent"). However, beliefs *in* the existence of an object can also be viewed as beliefs *about* the object, i.e., as beliefs linking the object to the concept of existence (e.g., "God exists"). Thus, without any loss in generality, beliefs in an object may be viewed as a special case of beliefs about the object, and the remainder of this book will be concerned only with beliefs about an object.

information about such a relationship is direct observation; that is, a person may perceive (via any of his sense modalities) that a given object has a certain attribute. For example, he may see or feel that a given table is round, he may taste or smell that a given glass of milk is sour, or he may see that a given person has dark skin. These direct experiences with a given object result in the formation of *descriptive beliefs* about that object. Since the validity of one's own senses is rarely questioned, these descriptive beliefs are, at least initially, held with maximal certainty. (Over time, forgetting may reduce belief strength.)

Clearly, however, an individual forms beliefs that go beyond directly observable relationships. For example, interaction with another person may lead to the formation of beliefs about such unobservable characteristics or dispositions as the person's honesty, friendliness, introversion, or intelligence. Bruner (1957) has suggested two ways in which an individual can go beyond observable events. First, he may make use of previously learned relationships. For instance, obese people are often assumed to be jolly; a person who is crying is assumed to be sad; and observation of smoke leads to the assumption of fire. A second way of going beyond observables involves the use of "formal coding systems." On the basis of the observations that Joe is taller than Ralph and Ralph is taller than Harry, a person may form the belief that Joe is taller than Harry—even though he has never observed Joe and Harry together. Formal coding systems thus refer to various rules of logic that allow the formation of beliefs about unobserved events.

Beliefs that go beyond directly observable events may be called *inferential beliefs*. The inferential beliefs in the examples above were based on prior descriptive beliefs. Thus the descriptive belief formed on the basis of the observation that "*O* is obese" produced the inference "*O* is jolly." Similarly, the descriptive beliefs "Joe is taller than Ralph" and "Ralph is taller than Harry" provided the basis for the inference that "Joe is taller than Harry." It should be clear, however, that inferential beliefs need not be based on descriptive beliefs but may instead be formed on the basis of prior inferences. For example, the inferential belief that *O* is jolly may lead to the further inference that *O* is happy. Although an inferential belief can thus be based on a prior inference, in the final analysis most inferences can be traced to descriptive beliefs.

It is worth noting that the distinction between descriptive and inferential beliefs is somewhat arbitrary. Many attributes of an object that appear to be direct observations cannot be directly perceived. For example, attributes such as round, sour, obese, happy, and dark are themselves concepts that have been acquired in the past. In Chapter 2 we pointed out that concept formation involves the conditioning of a common response or label to a set of discrete stimuli. The label "obese," for instance, is associated with certain physical dimensions of a person or animal. "Obesity" may be viewed as an attribute that is inferred from observation of certain characteristics of the object. Once such attributes or concepts have been well learned, however, a person tends to perceive the attribute or concept directly instead of perceiving a set of discrete stimuli. Thus one perceives trees, dogs, roundness, obesity, etc., and directly associates these concepts and attributes

in the formation of what were previously called descriptive beliefs. It is thus possible to view beliefs as representing a continuum from descriptive to inferential. At the descriptive end of the continuum, a person's beliefs are directly tied to the stimulus situation, and at the inferential end, beliefs are formed on the basis of these stimuli as well as residues of the person's past experiences; the continuum may be seen as involving minimal to maximal use of such experiential residues. For instance, a subject in an experiment is usually found to form beliefs about the race and sex of other participants, about the number of members in his group, about whether another person agreed or disagreed with him, etc. These are examples of relatively pure descriptive beliefs. Further along toward the inferential end of the continuum, McEwen and Greenberg (1969, 1970) varied the intensity of modifiers and verbs included in a communication. In one study, subjects were asked to rate the potency of the speaker's language and in another, the extremity of his position; in both cases, the appropriate inference was made. At the extreme inferential end of the continuum, beliefs may be almost entirely self-generated. For example, Miller's (1970) subjects were asked to "record their impressions" of a stranger by checking one member of each of 170 adjective pairs. The only information they had about the stranger was his or her photograph.

Many of our beliefs are formed neither on the basis of direct experience with the object of the belief nor by way of some inference process. Instead, we often accept information about some object provided by an outside source. Such sources include newspapers, books, magazines, radio and television, lecturers, friends, relatives, coworkers, etc. For example, we may read in the magazine *Time* that Paul Newman wears elevator shoes. On the basis of this information we may indeed form the belief that Paul Newman wears elevator shoes. Beliefs formed by accepting the information provided by an outside source may be termed *informational beliefs*. Although direct observations of an object-attribute relation will usually lead to the formation of a descriptive belief, outside information that links an object (e.g., Paul Newman) to an attribute (e.g., wears elevator shoes) may or may not lead to the formation of an informational belief. Many factors determine the degree to which information provided by an outside source will be accepted (see Chapter 11).

Whether or not the person forms an informational belief, exposure to information provided by an outside source will usually lead to the formation of a descriptive belief. That is, the person will come to believe that the source provided information concerning the relation between an object and some attribute. In the example above, he would have directly observed and formed the descriptive belief that "Time Magazine said that Paul Newman wears elevator shoes."

Later chapters will show the importance of this distinction between believing that an object (O) has an attribute (X) and believing that a source (S) provided information that O has the attribute X. The following schematic representation illustrates these two beliefs.

(O) is (X).
(S) said $(O$ is $X)$.

In the first case, belief formation involves the establishment of a link between *O* and *X*. In the second case, a source *S* makes the assertion that *O is X*; the *O-X* link itself is established by the source. In some instances formation of the descriptive belief "(*S*) said (*O* is *X*)" may lead to the informational belief "(*O*) is (*X*)" and in other instances it may not. Clearly, then, evidence that a person believes or "knows" that a given source asserted a relation between *O* and *X* cannot be taken as evidence that the person accepts the assertion itself, i.e., believes that *O* is related to *X*.

To summarize briefly, three different processes may underlie belief formation. First, a link between *O* and *X* may be actively established on the basis of direct observation (descriptive belief). Second, a link between *O* and *X* may be actively established through a process of inference from some other belief about *O* (inferential belief). Finally, a link between *O* and *X* may be established by some source, and this link may be accepted (informational belief).

We noted in Chapter 3 that a belief involves a link between an object and a content category (an attribute) and that belief strength refers to the subjective probability that the object is associated with the category in question. Strictly speaking, therefore, a belief is formed as soon as an object is linked to an attribute, irrespective of the subjective probability associated with the link. In theory, an individual may be viewed as having some subjective probability between 0 and 1 for all conceivable object-attribute links. Of course, he will be aware of only a limited number of such associations; that is, a person usually has a relatively small number of beliefs about any given object. Nevertheless, when presented with a novel object-attribute combination, he can indicate his subjective probability that the object has the attribute. Clearly, his response must in this case be an inference based on his prior beliefs. It follows that the mere presentation of an object and an attribute in combination (as in an opinion questionnaire) may result in the formation of inferential beliefs. To be sure, when the person has little information on which to base the inference, his subjective probability may be at chance level, indicating a high degree of uncertainty.

In this chapter we will discuss empirical research dealing with the formation of descriptive and inferential beliefs. Most research on the formation (and change) of informational beliefs has been conducted in investigations of persuasive communication and will be discussed in Chapter 11. Since the formation of descriptive beliefs has received relatively little attention, the present chapter will deal primarily with processes of inferential belief formation. Following a brief discussion of descriptive beliefs, we will turn to an examination of research related to the two bases of inferential belief formation identified by Bruner (1957). Thus we will first consider inferences based on perceived or learned relations between beliefs and then review research dealing with more formal models of the inference process. We will show that many different lines of investigation fall within one or the other of these two broad categories. For the most part, these investigations have not been viewed as falling within the attitude area. From our point of view, however, research on belief formation is of essential importance, both in its own

right and as the basis for an understanding of processes underlying the formation and change of attitudes, intentions, and behaviors. Our approach is essentially based on an information processing model, and a person's beliefs represent the information he has about himself and his social and physical environment. The first question of interest concerns the extent to which a person's descriptive beliefs formed on the basis of personal experience correspond to reality. Whether descriptive beliefs are veridical or inaccurate, we assume that inferences based on such beliefs are made in a systematic and predictable manner. That is, the information available to an individual is assumed to be processed in an internally consistent and orderly fashion, relatively uninfluenced by nonrational or dynamic forces.

DESCRIPTIVE BELIEFS

Manipulation Checks

Much of the information about the formation of descriptive beliefs in the attitude area comes to us in the form of incidental data collection or manipulation checks. In a typical experiment, subjects are exposed to a complex situation containing physical objects, persons, instruments, instructions, etc. Independent variables are manipulated by systematic variation in some of these stimuli. For example, some subjects may interact with a male experimenter, some with a female; the experimenter may reward or punish the subject; another person may agree with the subject's judgments; different subjects may be exposed to different communications, etc. Information about the formation of descriptive beliefs is obtained when the experimenter attempts to check the effectiveness of such manipulations by asking his subjects some direct questions about the manipulations. Subjects have been asked to indicate whether they had been rewarded or punished by the experimenter, whether another person had agreed or disagreed with their judgments, who the source of a communication was, etc.

Not only are subjects found to form beliefs on the basis of their observations; the bulk of the evidence indicates that the beliefs formed tend to accurately reflect what occurred in the situation—i.e., they tend to be *veridical.* Basically, beliefs assessed by manipulation checks represent tests of the subject's ability to recognize or recall events that occurred during the experiment. This question of memory has been investigated most intensively in studies on verbal learning.

Verbal Learning and Memory

Although research on verbal learning is probably as problematic as research in the attitude area (cf. Tulving and Madigan, 1970), some of the phenomena investigated by learning theorists have interesting implications for an understanding of belief formation. In Chapter 2 we showed that within the framework of learning theory, beliefs can be viewed as stimulus-response bonds. Any information con-

cerning the establishment of stimulus-response bonds may therefore have implications for belief formation. Thus interval between observation of object and attribute, number of object-attribute presentations, reinforcement of object-attribute associations, etc., may all influence belief strength.

It may prove useful, then, to examine some of the phenomena being investigated in the area of verbal learning and memory. Most investigators in the area (Bahrick, 1965; Postman, Jenkins, and Postman, 1948) have tended to view recognition and recall as indicants of the same learned stimulus-response relationship, although recognition is assumed to be a more sensitive measure in that it tends to show effects of learning where recall would not. After some practice in memorizing a list of words, for example, a person will usually be able to recognize more of them (when they are presented among other words) than he will be able to recall freely.[2] This implies that measures of recognition may lead to different conclusions concerning belief formation than will measures of recall. For example, Osterhouse and Brock (1970) found that distraction during a persuasive communication reduced recognition of persuasive arguments but had no significant effect on a measure of recall. Different results for recognition and recall were also reported by Zimbardo *et al.* (1970).

A phenomenon of considerable interest to the attitude area is the *serial position effect,* or the finding that recall of stimuli presented serially to the subject depends in part on their position in the sequence. Specifically, stimuli presented at the beginning and end of a list tend to be recalled better than stimuli appearing in the middle. Figure 5.1 illustrates the serial position effect in a typical learning

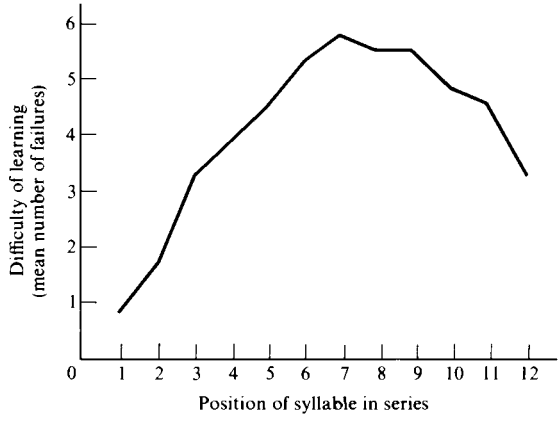

Fig. 5.1 Number of failures for each syllable position in the Hovland (1938) experiment. (Adapted from Hovland, Janis, and Kelley, p. 118.)

2. Recently, Adams and his associates (e.g., Adams and Bray, 1970; Adams, McIntyre, and Thorsheim, 1969; Adams, 1967) have suggested that different processes may underlie recall and recognition; the former is assumed to be based on a memory trace and the latter on a stimulus or response trace.

experiment, in which subjects were read a list of 12 three-letter syllables (Hovland, 1938). The implications of this phenomenon for the attitude area becomes apparent when a communication such as a persuasive message is viewed as a sequence of arguments to which a subject is exposed. The serial order effect suggests that subjects should be better able to recall what the communicator said at the beginning and end of his communication than what he said in the middle. A study by Jersild (1929) provided support for this argument (see Fig. 5.2). In that study, subjects read a series of 70 narrative biographical statements, and the pattern of recall exhibited the serial position effect.

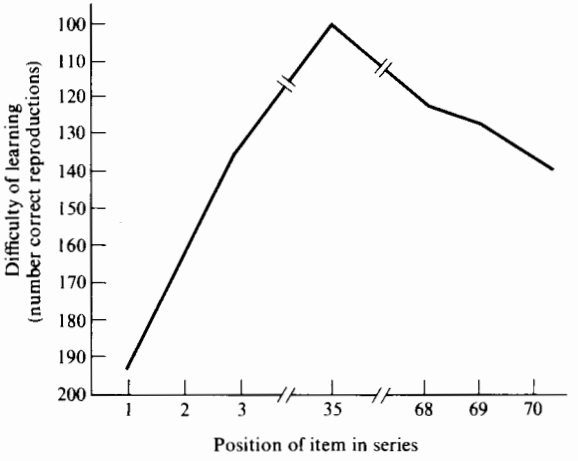

Fig. 5.2 Recall scores as a function of serial position of narrative items in the Jersild (1929) experiment. (Adapted from Hovland, Janis, and Kelley, p. 118.)

The serial position effect is found consistently for an ordered sequence of *homogeneous* verbal stimuli—nonsense syllables, meaningful words, or belief statements. However, the probability that stimuli in the middle of a sequence will be recalled can be increased by various manipulations. For example, when unfamiliar, novel, or unique stimuli that contrast with the remaining stimuli are inserted in the middle portion of a sequence, they are likely to be recalled.[3] A similar phenomenon occurs in short-term memory when subjects are asked to recall each individual stimulus in a list 10 or 20 seconds after its presentation. When homogeneous sets of stimuli are used, the percentage of subjects who correctly recall each succeeding stimulus tends to decline. As soon as stimuli of a

3. This is an example of the Von Restorff phenomenon. For a review, see Wallace (1965).

different kind are introduced, however, short-term memory tends to improve.[4] For example, Wickens (1970) has shown that shifting from words on one pole of the evaluative dimension (e.g., religious, success, nice, knowledge) to words on the opposite pole (e.g., hill, danger, lose, disease) results in a marked improvement in short-term memory. This effect can be seen in Fig. 5.3, which illustrates introduction of the novel evaluative stimulus on the fifth trial. Wickens obtained similar results with shifts along the activity and potency dimensions, shifts from high- to low-frequency words, and shifts from one class of nouns (e.g., birds) to another (e.g., occupations).

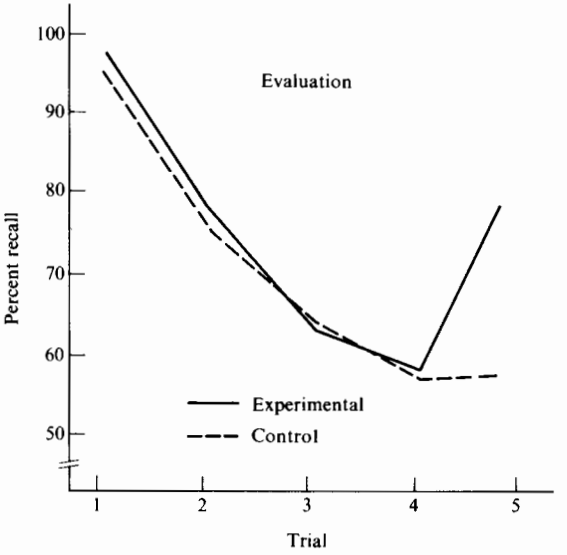

Fig. 5.3 The effect of shifting from one end to the other end of the Evaluative dimension. (Adapted from Wickens, 1970, p. 5.)

This phenomenon has potential implications for the *recency-primacy* issue in studies of impression formation and persuasive communication. A recency effect is obtained when information presented last has a stronger impact than information presented first, and the reverse is true for a primacy effect. In the context of the present discussion, the question is whether information presented first or last will be better recalled.[5] In a typical study, one side of an issue is presented and

4. These effects have been attributed to the operation of proactive inhibition and its release (Wickens, 1970) although increased rehearsal has been offered as an alternative explanation.

5. The recency-primacy issue is usually studied with respect to persuasion, i.e., *acceptance* of belief statements rather than their recall.

followed by a presentation of the opposite side, or a person is first described negatively and then positively (or vice versa). The research discussed above indicates that serial order position may be relatively unimportant for recall since the shifts from pro to con arguments or from positive to negative descriptions will tend to prevent deterioration of recall for information that occupies an intermediate position in the sequence of events.

Studies on verbal learning have also shown that meaningful, familiar stimuli are better recalled than meaningless, unfamiliar stimuli and that positively or negatively evaluated stimuli are better recalled than stimuli that are neutrally evaluated. The latter effect has also been found in the attitude area. McLaughlin (1970) described three persons in terms of 20 pieces of information; subjects were better able to recall information about the persons they liked and disliked than about the person for whom they had intermediate liking. Similar results were reported by Lott, Lott, and Walsh (1970), and somewhat related to these findings, Johnson, Middleton, and Tajfel (1970) found that children had more factual knowledge about liked and disliked than about neutral countries.

Other phenomena related to verbal learning and memory are potentially relevant to belief formation and change. A complete discussion of this research, however, is beyond the scope of this book. The primary aim of the discussion above is to draw the reader's attention to a relevant body of literature that has often been ignored by researchers in the attitude area.[6] The main interest of attitude researchers has been in the recognition or recall of the contents of a persuasive message.

Recall of Message Content

Whenever a subject is exposed to a communication from some source S which links an object O to an attribute or concept X, he is likely to form the belief that S said O *is* X. That is, the subject will usually be able to recognize or recall what the communication said. However, he may or may not accept (i.e., believe) the communication's content. For present purposes we are interested only in the subject's ability to recognize or recall events that occurred during an experiment, i.e., in his belief that S said O *is* X. Many studies have attempted to assess these beliefs by giving tests of recognition, recall, reception, or learning. Each of these methods involves one of two procedures: Either the subjects are asked to reproduce the content of the communication (*recall*), or they are given a set of statements and are asked to indicate which of the statements were part of the message (*recognition*). Since most recognition tests are presented in a multiple-choice or true-false format, the subject may respond by selecting either a statement that he believes was part of the communication (i.e., recognition) or a statement that represents his own belief (i.e., acceptance). Unfortunately, many research reports

6. Interested readers are directed to Adams (1967) and Tulving and Madigan (1970) for reviews of some of the literature in verbal learning and memory.

describe the measure of reception or learning that was used without an explicit indication as to the judgment the subjects were required to make. If subjects are given no specific instructions, their responses may reflect either recognition or acceptance.

Thus it is quite possible that many studies employing some measure of reception or learning have confounded recognition and acceptance; this may in part account for some of the conflicting findings about effects of different variables on recognition of message content. When only studies with measures of recall or clear measures of recognition are considered, however, results appear to be consistent with findings concerning verbal learning and memory.

One issue that has attracted repeated attention is the degree to which a person's own beliefs or attitudes affect his recall or recognition of belief statements contained in a communication. Perhaps on the basis of psychodynamic theory, it has often been assumed that, as a defensive reaction, individuals will reduce their attention to or comprehension of unpleasant, inconsistent, or threatening material. Thus it has been hypothesized that a person will be better able to recall information consistent with his own beliefs or attitudes than inconsistent information.

One of the first studies to clearly demonstrate this effect was conducted by Levine and Murphy (1943) who used five pro- and five anti-Communist college students as their subjects. Each subject participated in 10 sessions; in each of the first five sessions (learning phase) the subject read two prose passages, one of which was strongly anti-Communist, the other moderately pro-Communist. Subjects were asked to read the first passage twice and 15 minutes later to reproduce it as accurately as possible. The procedure was then repeated for the second passage. In the last five sessions (forgetting phase), only the recall task was performed. The results indicated that in all 10 sessions, exact reproduction of "idea groups" in a given passage was greater for subjects whose attitudes were consistent with that passage than for subjects whose attitudes were inconsistent with it. That is, pro-Communist subjects correctly recalled more of the pro-Communist passage than did anti-Communist subjects, and the reverse was true with respect to the anti-Communist passage.[7]

These findings have frequently been cited as evidence that a person's own attitude influences his recall of belief statements contained in a communication. However, many other studies performed before and after the Levine and Murphy experiment did not support this hypothesis (Watson and Hartman, 1939; A. L. Edwards, 1941; Jones and Aneshansel, 1956; Brigham and Cook, 1969; Malpass, 1969). In fact, most investigations have found little or no effect of a person's own beliefs or attitudes on his recall of belief statements to which he has been exposed. Other variables are usually found to play a more important role in deter-

7. However, all subjects recalled more of the pro-Communist than of the anti-Communist passage, a result that is perhaps attributable to the different contents and styles (e.g., use of extreme or unusual words, etc.) of the passages.

mining recall of message content. Brigham and Cook (1969), for example, failed to replicate the Levine and Murphy study. Subjects were exposed to a "transcript" of a discussion on integration that contained eight prointegration and eight anti-integration statements. Four statements in each set were considered plausible, the other four implausible. Subjects were selected on the basis of a multifactor racial attitude inventory (Woodmansee and Cook, 1967); data for subjects with the 22 highest and 22 lowest racial attitude scores were analyzed. Measures of free recall and recognition were obtained. There was no evidence that subjects recalled or recognized more statements consistent with their own attitudes than inconsistent statements. Jones and Kohler (1958), however, had found some support for the argument that the Levine and Murphy finding obtains only with plausible statements and that subjects recall more inconsistent than consistent statements when they are implausible. Although Brigham and Cook found better recall for implausible than for plausible statements, the Jones and Kohler findings were not replicated; that is to say, implausible statements were better recalled irrespective of the subject's own attitude or the position expressed in the statements.

In conclusion, there appears to be little evidence for the hypothesis that belief statements consistent with a person's own position are recalled or recognized better than belief statements inconsistent with his position. Indeed, as the research on verbal learning and memory suggests, many factors other than supportiveness tend to affect the degree to which information is recalled. Thus arguments differing in their supportiveness may also vary in terms of their plausibility, evaluative polarity, novelty, or familiarity, and such factors are likely to influence recognition and recall. Further, it may sometimes be to the person's advantage to recall certain arguments regardless of their supportiveness. For example, Jones and Aneshansel (1956) found that recall for nonsupportive arguments could be increased by telling subjects that they would be able to use those arguments in a later task. Such factors, therefore, have to be taken into consideration when selective recall is investigated.

Factors of this type are less likely to be involved in investigations of the effects of beliefs about the source or attitudes toward the source on recall of a communication from the source. In these studies, all subjects are exposed to the same communication. A typical experiment manipulates the expertise, trustworthiness, or attractiveness of the source, and the effects of these manipulations on recognition and recall of message content can be studied. The evidence overwhelmingly indicates that neither beliefs about the communicator's credibility nor attitudes toward the communicator have any appreciable effects on a subject's recognition or recall of the contents of his communication (Hovland, Janis, and Kelley, 1953; Johnson and Scileppi, 1969; Singer, 1969). For example, Johnson and Scileppi's subjects read a communication atttributed to one of two sources. One source was described as "a medical authority who is recognized as an expert on the issue of chest X-rays and tuberculosis" and the other source as "a medical quack who had served a prison term for medical fraud and who knew nothing substantial about the issue, but had written the communication for a magazine catering to sensationalism"

(p. 32). This extreme manipulation of source credibility had no significant effects on a measure of recall of message content.

A third question concerns the effects on recognition or recall of stable individual-difference variables, such as sex, self-esteem, intelligence, chronic anxiety, dogmatism, authoritarianism, etc. The limited evidence available on this question suggests that variables of this type have little or no systematic effect on recognition or recall of the contents of a communication.[8] There is little evidence, then, that a person's beliefs about or attitude toward a source or the topic of his communication influence recognition or recall of message content or that stable individual-difference variables have any consistent effects. Studies in verbal learning and memory suggest, however, that several other factors may influence such recall. For instance, interference during the learning stage and motivation have been found to affect recognition and recall. Similarly, in the attitude area recall has been influenced by drug-produced arousal (Friedman, Buck, and Allen, 1970) and by some but not all types of distractors or manipulations of acute fear or anxiety (e.g., Haaland and Venkatesan, 1968; Zimbardo *et al.*, 1970; McArdle, 1972). Some of these factors will be considered in our discussion of communication and persuasion in Chapter 11.

Recall of Contingencies

We have tried to show that descriptive beliefs are formed on the basis of direct observation. A final line of research providing support for this notion comes from studies dealing with the role of awareness in verbal conditioning. In our discussion of reactive effects (Chapter 4) we pointed out that subjects tend to form beliefs about the experimental situation, including beliefs about the consequences of their own behavior and about contingencies between other events. In fact, it has been found that verbal "conditioning" does not occur unless subjects are aware of these contingencies (Dulany, 1961, 1968; DeNike and Leibovitz, 1969; Uleman, 1971; Page, 1969). Some studies have employed a classical conditioning paradigm in which neutral stimuli are consistently paired either with positively or negatively evaluated stimuli. Changes in evaluation of the neutral stimuli are found only when subjects are aware of these consistent pairings. Other studies have used an operant conditioning paradigm in which certain responses made by the subject are reinforced by the experimenter. Again, increments in response strength are observed only for subjects aware of the reinforcement contingencies. A more detailed discussion of research on the conditioning of attitudes will be presented in Chapter 6; at present we note only that observation of events occurring in some situation leads to the formation of beliefs.

Most research has been satisfied with demonstrating that such beliefs are

8. Effects of some of these variables (e.g., intelligence) may be somewhat obscured because of the relatively narrow range of scores among subjects sampled from a student population.

indeed formed, but some studies have also examined factors that may influence the formation of beliefs about contingencies. For example, Uleman and Vanden-Bos (1971) reported that the formation of the correct contingency hypotheses in an operant conditioning paradigm was more likely to occur with a short than a long delay of reinforcement. The reinforcer in this study was a buzzer, and the duration of buzzing was also manipulated. The effect of this manipulation interacted with delay of reinforcement in influencing the formation of correct contingency hypotheses. Thus, with a three-second delay, contingency awareness increased with duration of buzzing, but with a five-second delay, variations in duration had no effect on contingency awareness. Finally, in this study as well as in a study by Dulany (1968), instructions as to the meaning of the reinforcing event were found to influence the formation of beliefs about contingencies.

Schedules of reinforcement may also be expected to have some effect on the formation of beliefs about contingencies. For example, subjects reinforced on a 100 percent schedule should be more likely to form the hypothesis in question than subjects on a 50 percent or random schedule. Some indirect evidence comes from studies which show that subjects' predictions of the occurrence of future events tend to correspond to the relative frequency with which such events have previously occurred (e.g., Humphreys, 1939; Estes, 1964).

To summarize, it is evident that people tend to form beliefs that reflect the things they experience, and under most circumstances, the beliefs formed tend to be fairly veridical. Nevertheless beliefs are sometimes found to be nonveridical. This is especially likely when events occur unexpectedly or when the person does not have sufficient time to observe the event carefully. For example, witnesses of a crime or an accident often form different beliefs about the events that occurred. Other evidence for occasional nonveridicality in descriptive belief formation comes from studies on optical illusions. The illusions occur in situations where stimuli are arranged in such a way as to produce incorrect perception.

With the exception of illusions, however, there appears to be little systematic bias in descriptive belief formation. That is, there is practically no evidence to support the notion that a person's own beliefs, attitudes, or personality characteristics have any systematic effects on descriptive beliefs based on direct observation. This should not be taken to mean that all types of beliefs are unaffected by such variables; on the contrary, some of them play an important role in the formation of beliefs that go beyond direct observation; i.e., they will tend to affect the inference process.

INFERENTIAL BELIEFS

The most characteristic feature of inferences is that in addition to the stimulus situation, the individual uses residues of past experience to make his judgments. Consider, for example, a person who has a negative attitude toward the People's Republic of China and who positively values freedom of religion. Imagine further that this person has never received any direct information about the degree of

religious freedom in China and that he has in fact never considered the link between these two concepts. If he were now asked to indicate his subjective probability that China has religious freedom, he would probably have little difficulty in making a judgment. Clearly, his judgment would represent an inference rather than a descriptive belief. This part of the present chapter deals with the processes underlying such inferential belief formation.

One possible basis for the formation of inferential beliefs was discussed in Chapter 2, namely, Heider's (1944, 1958) notions concerning causal attribution and balance. Heider suggested that two elements will be perceived as forming a unit relation when both elements have the same dynamic character, i.e., when both are positively or negatively evaluated. When one element is evaluated positively and the other negatively, the person will perceive no relation between the elements; i.e., they will tend to be segregated. Similarly, the reader should recall that in a triadic configuration, $p\mathrm{L}o$ and $p\mathrm{L}x$ imply $o\mathrm{U}x$, and $p\overline{\mathrm{L}}o$ and $p\mathrm{L}x$ imply $o\overline{\mathrm{U}}x$. The latter configuration is a balanced triad that is applicable to our example above. Since the person in question has a negative attitude toward China and a positive attitude toward freedom of religion, he should, according to Heider's theorizing, form the inferential belief that China does not have freedom of religion. His subjective probability that China has religious freedom should therefore be low.

This judgment is representative of the kinds of judgments that subjects are typically asked to make when they respond to standard attitude scales. As mentioned in Chapter 3, most standard scaling methods are designed to select items that have a given *p-x* link; agreement or disagreement with the item is then assumed to be primarily a function of the *p-o* link, i.e., of the person's attitude toward the object under consideration. In fact, many indirect and almost all disguised measures of attitude are based on the assumption that responses to certain statements are primarily determined by the subject's own attitude.

These considerations imply that evaluation may provide one basis for inferential belief formation. In its most general form, this principle suggests that inferences follow along the lines of *evaluative consistency*. A person is expected to infer that liked objects have positive attributes and that disliked objects have negative attributes. Conversely, he is not expected to infer that liked objects have negative attributes or that disliked objects have positive attributes.

Clearly, however, the person may form inferential beliefs as a result of other processes. In the example above, the subjective probability judgment may be the conclusion based on syllogistic reasoning. Thus the person might hold the following two beliefs.

1. The People's Republic of China is a Communist country.
2. Communist countries do not have religious freedom.

On the basis of these beliefs he might form the following inference.

3. China does not have religious freedom.

Note that although the person has formed the same inferential belief, the underlying process did not involve evaluative consistency; instead, it was based on considerations of relations between different beliefs. Although the inference process need not follow the rules of *formal* logic, it tends to be internally consistent. The consistency in this case refers to relations between cognitions or beliefs. To contrast this type of consistency with evaluative consistency as a basis for inferential belief formation, we will refer to it as *probabilistic consistency.*

These considerations suggest that an inferential belief linking an object and an attribute in an evaluatively consistent manner may actually be based on probabilistic rather than evaluative consistency. A simple example is given by the inferential belief that "hypochondriacs are anxious." Since both "hypochondriac" and "anxious" are negatively evaluated, the belief exhibits evaluative consistency, and it is possible that this evaluative consistency led to the inference. More likely, however, this belief is formed in a more direct probabilistic fashion since hypochondria is usually viewed as related to anxiety. (That is, on the basis of prior learning, the conditional probability that a person is anxious if he is a hypochondriac will tend to be high.)

A more interesting example is given by a person's inferential belief that "Khrushchev was immature." In terms of evaluative consistency this inference would be made by individuals who disliked Khrushchev as well as immaturity; in terms of probabilistic consistency, this inference could be a function of the belief that Khrushchev took off his shoe and banged it on the table in the United Nations and the belief that people who behave in this way are immature.[9]

Inferences Based on Relations between Beliefs

The examples above should make clear that it will often be impossible to distinguish between inferences based on evaluative consistency and inferences based on probabilistic consistency. The confounding between evaluative and probabilistic consistency was recognized in research on trait inferences (Peabody, 1967). Although much of this research has been conducted outside the attitude area, it is obviously relevant for our present discussion of inferential belief formation.

Trait Inferences

Initial impetus to much of the work on trait inferences came from Asch's (1946) studies of impression formation. In one experiment, subjects were read one of two lists of seven personality traits that were supposedly descriptions of a real person; the lists were identical except for the fourth trait in the list, which was either *warm* or *cold.* The two lists used were: intelligent, skillful, industrious,

9. Incidentally, the belief about Khrushchev's behavior may be in part responsible for a person's negative attitude toward Khrushchev as well as for the inference that Khrushchev was immature. This newly formed inferential belief will also contribute negatively to the person's attitude toward Khrushchev.

(warm or cold), determined, practical, cautious. After hearing the list, subjects were asked to write brief descriptions of the person the traits brought to mind, and to indicate for each of 18 pairs of different bipolar traits the term that best fitted the impression they had formed. The positive member of each trait pair is shown in Table 5.1.

Not only did the two lists of traits produce qualitatively different written impressions, but there were huge quantitative differences with respect to some of the items on the checklist. For example, when the description included the trait *warm,* most subjects perceived the person involved to be generous, humorous, sociable, popular, and good-natured. Inclusion of the trait *cold* in the stimulus list led not to these generally favorable impressions but to their bipolar opposites.

In a second experiment, Asch used the same stimulus lists, except that the traits *polite* and *blunt* were substituted for *warm* and *cold.* Differences in impressions produced by the two new lists were relatively minor, and Asch concluded that in the context of the stimulus lists employed, the *warm-cold* variable was more central to impression formation than the *polite-blunt* variable. The results of these two experiments are given in Table 5.1. In still other experiments, the traits *warm* and *cold* were embedded within new sets of stimulus traits, and their effects in some of these lists tended to be relatively small; that is, *warm* and *cold,* like *polite* and *blunt* in the previous study, operated as peripheral rather than central traits.

In another series of experiments, Asch varied the order of traits describing a stimulus person. Specifically, a given set of traits was read to the subjects in one of two orders. For example, in one experiment the lists were read in the following orders.

Order A: intelligent-industrious-impulsive-critical-stubborn-envious

Order B: envious-stubborn-critical-impulsive-industrious-intelligent

As Table 5.1 shows (Experiment 3), subjects receiving Order A were more likely than those receiving Order B to infer that the stimulus person was happy, good-natured, good-looking, and restrained. Asch reported somewhat weaker order effects when he used another set of traits in the following two orders (Experiment 4).

Order A: intelligent-skillful-industrious-determined-practical-cautious-evasive

Order B: evasive-cautious-practical-determined-industrious-skillful-intelligent

Asch concluded that these studies demonstrated a primacy effect in impression formation. In Experiment 3, the traits on the list were ordered in terms of their evaluation such that Order A gradually shifted from positive to negative traits, and the reverse was true for Order B. In Experiment 4, "unlike the preceding series, there is no gradual change in the merit of the given characteristics, but

Table 5.1 Results of Asch's Impression Formation Experiments

	Centrality effects				Order effects			
	Experiment 1		Experiment 2		Experiment 3		Experiment 4	
	Warm ($N = 90$)	Cold ($N = 76$)	Polite ($N = 20$)	Blunt ($N = 26$)	Order A ($N = 34$)	Order B ($N = 24$)	Order A ($N = 46$)	Order B ($N = 53$)
1. generous	91[a]	8†	56	58	24	10	42	23
2. wise	65	25†	30	50	18	17	35	19
3. happy	90	34†	75	65	32	5*	51	49
4. good-natured	94	17†	87	56	18	0*	54	37
5. humorous	77	13†	71	48	52	21*	53	29
6. sociable	91	38†	83	68	56	27	50	48
7. popular	84	28†	94	56	35	14	44	39
8. reliable	94	99	95	100	84	91	96	94
9. important	88	99	94	96	85	90	77	89
10. humane	86	31†	59	77	36	21	49	46
11. good-looking	77	69	93	79	74	35*	59	53
12. persistent	100	97	100	100	82	87	94	100
13. serious	100	99	100	100	97	100	44	100*
14. restrained	77	89	82	77	64	9†	91	91
15. altruistic	69	18†	29	46	6	5	32	25
16. imaginative	51	19†	33	31	26	14	37	16
17. strong	98	95	100	100	94	73	74	96
18. honest	98	94	87	100	80	79	66	81

[a] Percentage of subjects attributing positive trait to the stimulus person
* $p < .05$
† $p < .01$

rather the abrupt introduction at the end (or at the beginning) of a highly dubious trait" (Asch, 1946, p. 272). The primacy effect refers to the finding that the lists starting with favorable traits (Order A) led to a greater number of favorable inferences (see Table 5.1). Thus, in Experiment 3, Order A led to more favorable inferences than Order B on 14 of the 18 response traits; in Experiment 4, this difference was observed on 12 of the 18 response traits. However, the order effect is usually quite small when any given inference is considered. Indeed, a statistical analysis of Asch's data reveals only five significant differences in Experiment 3 and one significant difference in Experiment 4. Further, the latter difference indicates a recency rather than a primacy effect.[10]

More on the basis of the subjects' qualitative descriptions of the stimulus person than on these quantitative data, Asch suggested a "change in meaning" explanation for both the centrality and order effects. He argued that within a given set, some traits play a more central role than others. "The whole system of relations determines which will become central. These [central traits] set the direction for the further view of the person and for the concretization of the dependent traits." (Asch, 1946, p. 284) Thus traits appearing first in a list often become central and direct the overall impression by influencing the meaning of the subsequent traits. However, Asch pointed out that "it is not the sheer temporal position of the item which is important. . . . The effects of primacy should be abolished —or reversed—if it does not stand in a fitting relation to the succeeding qualities, or if a certain quality stands out as central despite its position. The latter was clearly the case for the quality 'warm-cold' in Experiment 1 [see Table 5.1] which, though occupying a middle position ranked comparatively high." (Asch, 1946, p. 272) Further, when placed in a different context, the *warm-cold* variable appeared to occupy a less central position, and according to Asch, its meaning had changed.

Focusing on either the order effect or the centrality effect, other investigators have offered alternatives to Asch's directed impression or change in meaning explanation. To account for order effects in impression formation, it has been suggested (e.g., N. H. Anderson, 1968a; Anderson and Hubert, 1963; Anderson and Barrios, 1961) that the weight or importance of a given trait varies according to its position while its meaning remains constant. Specifically, a primacy effect would result if smaller weights were placed on each succeeding trait in the process of impression formation. Research concerning this question will be reviewed in Chapter 6 since this research has dealt almost exclusively with the formation of attitudes rather than beliefs. Indeed, Asch's experiments may be the

10. It is noteworthy that in the context of impression formation, primacy and recency effects must be considered with reference to some dimension common to stimulus and response traits. Asch defined his primacy effect in terms of an evaluative dimension. Although other dimensions could be considered, we shall see that virtually all research on order effects in impression formation has been concerned with the evaluative dimension or liking for the stimulus person.

only investigation of order effects in belief formation. As indicated earlier, a secondary analysis of his data showed few significant effects of order on inferential beliefs, and further, those results which were significant indicated both primacy and recency effects.

With respect to the question of centrality, several investigators (e.g., Wishner, 1960; Rosenberg, Nelson, and Vivekananthan, 1968) have proposed an explanation which revolves around the perceived relations between traits. Consider, for example, any set of 20 traits. For any pair in the set, subjects can be asked to indicate the likelihood that a person will have trait x if he has trait y. These conditional probabilities (that is, $P(x|y)$, the probability of having trait x, given trait y) can be obtained for all possible pairs of traits, and the matrix of 20×20 conditional probabilities represents a complete description of the perceived relations between traits. Alternatively, it is possible to have subjects rate one or more stimulus persons on the 20 traits. (For example, they could be asked to indicate the probability that the stimulus person has each trait.) These ratings can be used to compute an intercorrelation matrix between all possible pairs of traits, which would again represent a complete description of the perceived relations among traits; the higher the correlation, the greater the perceived relationship.

It may be argued that the introduction of a given trait into a stimulus list will affect responses to an adjective checklist to the extent that the trait in question (1) is *not* related to the other traits in the stimulus list and (2) *is* related to the traits in the response list. An early study by Bruner, Shapiro, and Tagiuri (1958) showed that perceived relations among individual traits could be used to make predictions concerning inferences based on two or three traits in combination. For example, one group of subjects indicated the degree to which an *intelligent* person would be likely to have 59 other traits, such as *aggressive, unreliable, submissive,* etc. A second group of subjects made the same inferences for a person who was *inconsiderate,* and a third group for a person who was both *intelligent* and *inconsiderate.* Generally speaking, the judgments made by subjects in the third group could be predicted from the judgments made by subjects in the first two groups.[11]

This finding suggests that in the Asch studies, impression formation may have been a function of the perceived relations among the stimulus traits and the response traits. Wishner (1960) obtained a matrix of intercorrelations for all stimulus and response traits used in the Asch study. The results provided considerable support for the argument that Asch's findings could be interpreted in terms of the relations within stimulus traits as well as between stimulus and response traits. That is, when *warm* and *cold* were "central" traits, they tended to have low cor-

11. Unfortunately, the investigators did not take full advantage of the available probabilistic data, relying instead on a rather gross distinction as to whether or not the majority of subjects in a given group perceived some degree of likelihood that one trait implied another. Thus no firm conclusions can be drawn on the basis of this study; it is mentioned here primarily for its pioneering role in using conditional probability measures.

relations with the other traits in the stimulus list and to correlate highly with those traits in the response list that were most influenced by the manipulation. Rosenberg, Nelson, and Vivekananthan (1968) also reported evidence in support of this argument. Using a multidimensional scaling procedure, these authors found that two or three factors could account for the relationships among the traits used in Asch's study, and they argued that Asch's results could be predicted from a knowledge of the loadings of the traits on these dimensions.

Much of the research on trait inferences has been concerned with describing the interrelationships among traits. Perhaps in response to Bruner and Tagiuri's (1954) argument that perceived interrelations among traits represent naive or common-sense "theories of personality," investigators turned their attention to the dimensions underlying perceived trait relations in order to identify the basic structure of these *implicit theories of personality*. Factor analysis and multidimensional scaling procedures have been employed in these investigations (e.g., Cattell, 1946; Tupes and Christal, 1961; Norman, 1963). An excellent review of this body of literature can be found in J. S. Wiggins (1973, Chapter 8).

Many of these studies are concerned with five major orthogonal dimensions that appear to account for much of the variance in trait inferences. Table 5.2 presents the results of a factor analysis based on intercorrelations of 20 bipolar adjective scales. These data are based on ratings of fraternity members by their peers (i.e., other fraternity members) on the 20 bipolar scales (Norman, 1963). The scales were selected to represent the five factors identified by Tupes and Christal (1961) in a study based on Cattell's (1946) pioneering work in this area. The same factor structure emerged in Norman's analysis (see Table 5.2). Many additional studies using the same set of scales have replicated the basic factor structure. The same factor structure emerged whether the factor analysis was performed on ratings of the stimulus persons on the 20 scales (Norman, 1963), on judgments of similarity between pairs of traits (D'Andrade, 1965), or on ratings of conditional probabilities between pairs of traits (Hakel, 1969). Further, the factor structure was replicated for ratings of complete strangers (Passini and Norman, 1966) as well as close acquaintances (Norman, 1963).[12]

It appears, then, that in the natural usage of trait descriptions, certain traits tend to form a consistent pattern, and the implicit structure of trait relations closely parallels the structure obtained when real persons are judged. Thus, once a given trait has been assigned to a person on any basis, "the remaining ratings

12. However, there was greater consensus among subjects rating a common acquaintance than among subjects rating a stranger. It has been argued (J. S. Wiggins, 1973) that when a subject rates a stranger, he is forced to rely on "superficial characteristics," such as dress, demeanor, physical size, etc., to infer some personality trait. Once a given rating has been made on this basis, the remaining traits are relatively fixed by the existing implicative structure among trait terms. Since subjects may make different initial inferences, there will be relatively little agreement in their descriptions of a stranger. For a common acquaintance, however, inferences and overall descriptions are likely to coincide.

Table 5.2 Factor Matrix of Peer Ratings in a Fraternity Sample (from Norman, 1963, p. 579)

Scales	Factors I	II	III	IV	V
Extraversion					
Talkative–Silent	**90**	02	−02	04	−00
Frank, Open–Secretive	**78**	−08	07	−03	07
Adventurous–Cautious	**78**	15	−20	32	01
Sociable–Reclusive	**86**	01	−18	−01	−02
Agreeableness					
Goodnatured–Irritable	17	**80**	17	12	07
Not jealous–Jealous	−10	**64**	20	49	07
Mild, Gentle–Headstrong	−20	**80**	27	19	10
Cooperative–Negativistic	33	**74**	28	13	11
Conscientiousness					
Fussy, Tidy–Careless	−33	−08	**66**	−35	20
Responsible–Undependable	−03	32	**86**	08	18
Scrupulous–Unscrupulous	−30	44	**68**	−02	20
Persevering–Quitting, Fickle	−05	28	**74**	12	27
Emotional stability					
Poised–Nervous, Tense	01	56	15	**61**	05
Calm–Anxious	06	21	−10	**82**	−07
Composed–Excitable	13	06	16	**71**	24
Not hypochondriacal–Hypochondriacal	21	27	−00	**65**	−09
Culture					
Artistically Sensitive–Insensitive	−04	08	39	−10	**75**
Intellectual–Unreflective, Narrow	−04	05	47	04	**74**
Polished, Refined–Crude, Boorish	15	25	53	16	**46**
Imaginative–Simple, Direct	12	19	03	10	**68**

are relatively 'fixed' by the implicative structure that exists among trait terms" (J. S. Wiggins, 1973). As Table 5.2 shows, if a stimulus person is perceived as *sociable,* he is also likely to be seen as *talkative, frank, open,* and *adventurous.* If he is considered *calm,* he is also likely to be perceived as *poised, composed,* and *not hypochondriacal.* These consistently emerging factors are generally viewed as representing perceived dimensions of personality.

Note, however, that the procedure whereby these factors were identified strongly resembles the procedure used by Osgood and his associates (Osgood, Suci and Tannenbaum, 1957) to identify the basic dimensions of meaning. That is to say, in both lines of research the investigator starts with a large number of adjec-

tives or bipolar adjective scales against which various stimuli are judged. Osgood's subjects rated stimuli from various domains (e.g., persons, animals, values, inanimate objects, etc.), whereas investigators studying "implicit theories of personality" used only persons as stimulus concepts. In both approaches, however, dimensions underlying these ratings were identified by factor analysis or multidimensional scaling. Perhaps it is not surprising, therefore, that many investigators (e.g., D. R. Peterson, 1965; Hallworth, 1965; Becker, 1960; Burke and Bennis, 1961) began to argue that the perceived dimensions of personality identified in research on trait inferences are "nothing but" general dimensions of meaning as applied to judgments of people. Thus Peterson (1965) argued that

> The invariant "personality" dimensions discussed above [neuroticism and extraversion] are rather easily construed as topical variants of more general ways of attributing meaning to objects, in this case human objects. "Adjustment" is good, "neuroticism" is bad. Extraversion "means" strong and active. (p. 57)

The diverging interpretations of the factors emerging in studies of trait inferences have led to the distinction made earlier between evaluative and probabilistic consistency in inferential belief formation. The central question is whether two traits that are perceived to be related are seen as related because they are evaluatively similar to each other or because they share descriptive or semantic similarity. For example, when a person who is judged to be *trustworthy* is also judged to be *cooperative,* the perceived relation between the two traits may be due to their common positive evaluative meaning or to their descriptive similarity in terms of dependability or stability.

As stated by Peabody (1967, p. 2),

> The problem is that a judgment is commonly at the same time both an estimate of the factual situation and an evaluation. For example, consider such contrasts as "kind-cruel" or "cautious-rash." The judgment that a person or action is "kind" or "rash" combines a descriptive aspect (e.g., that the action helps others, or involves very large risks) and an evaluative aspect (e.g., that the action is desirable or undesirable). In a single trait term (and its opposite) the two aspects are confounded since they always combine in the same way (e.g., helping as desirable, hurting as undesirable).

Thus Peabody argued that evaluative and descriptive similarity have been confounded in studies on trait inferences. In other words, it is not clear whether the formation of inferential beliefs is based on evaluative or probabilistic consistency.

To investigate this problem, Peabody devised a paradigm that attempts to separate evaluative and probabilistic consistency by orthogonally manipulating the evaluative and descriptive similarity of adjective pairs. Consider, for example, the positively evaluated trait *cautious* and an inference to the dimension defined by the endpoints *bold* (positive) and *timid* (negative). Peabody argued that if trait inferences were based on evaluative similarity, a person who was described as *cautious* should also be judged as *bold.* But if inferences followed descriptive similarity, he should be judged as *timid.*

Peabody constructed sets of four adjectives that varied in terms of evaluative and descriptive similarity. Table 5.3 presents four of the sets used by Peabody. The numbers in front of each trait represent the mean evaluation of the trait based on evaluative semantic differential ratings. Note that within each set of traits, two are positively evaluated (e.g., *cautious* and *bold*) and two are negatively evaluated (e.g., *timid* and *rash*). Further, each positive trait is descriptively similar to a negative trait (e.g., *cautious* is descriptively similar to *timid,* and *bold* is descriptively similar to *rash*). After constructing these sets, Peabody obtained judgments using items such as:

CAUTIOUS

bold |——|——|——|——|——|——|——| timid

Subjects were instructed to "assume a person with the characteristic given in capital letters. On the scale immediately beneath, you are to judge how likely it is that this person has one or the other of the traits given by the scale" (Peabody, 1967, p. 5).

Within each set of traits, four critical judgments of this kind are possible. For the first set in Table 5.3, the following *critical* judgments were obtained.

CAUTIOUS:	bold–timid
TIMID:	cautious–rash
BOLD:	cautious–rash
RASH:	bold–timid

In each item, the capitalized trait is descriptively similar to one member of the associated adjective pair and evaluatively similar to the other member of the pair.

In addition, each trait (e.g., CAUTIOUS) was rated on bipolar scales formed from traits in the other sets (e.g., on *thrifty–extravagant*). These *"unselected"* ratings are considered to represent typical trait inferences where evaluative and

Table 5.3 Sets of Trait Terms Differing in Evaluative Meaning (from Peabody, 1967)

+ .9 Cautious	+1.1 Bold
−1.1 Timid	−1.2 Rash
+1.7 Self-controlled	+1.1 Uninhibited
−1.4 Inhibited	− .3 Impulsive
+ .9 Thrifty	+1.8 Generous
−2.0 Stingy	− .8 Extravagant
+ .5 Skeptical	+1.8 Generous
−1.4 Distrustful	−1.4 Gullible

descriptive similarity are confounded. That is, it could be argued that the trait *cautious* is both descriptively and evaluatively similar to the trait *thrifty*. In most cases, however, the problem is not one of confounding evaluative and descriptive similarity but rather the fact that descriptive similarity may be involved but cannot be readily identified. For instance, when the trait CAUTIOUS is judged against *tense–relaxed,* it is not clear which, if either, of the two adjectives is descriptively more similar to cautious. Thus the critical judgments constituted attempts to separate evaluative and descriptive similarity, whereas these two aspects were confounded in the unselected judgments. Peabody's subjects rated 90 traits on 40 scales, thus making 3600 judgments. Of these judgments, 80 involved rating a trait that was also one of the bipolar adjectives of the scale; these judgments were excluded from the analyses. Of the remaining 3520 judgments, 70 were critical and 3450 were unselected.

With respect to unselected judgments, 67 percent followed evaluative similarity. Thus, when descriptive similarity is either ambiguous or confounded with evaluative similarity, the majority of judgments may be viewed as exhibiting evaluative similarity. In contrast, all 70 critical judgments followed descriptive similarity. Peabody therefore concluded that when evaluative and descriptive similarity are separated, inferences tend to follow lines of descriptive similarity.[13]

Using Peabody's procedure, Felipe (1970) replicated these findings. Of the 136 unselected judgments in this replication, 73 percent exhibited evaluative similarity. Of 32 critical judgments, however, 75 percent followed descriptive similarity. In addition, Felipe constructed two different types of response scales from the basic sets of four personality traits. The first type contrasted traits that were descriptively similar but differed in evaluation (e.g., *stingy–thrifty; extravagant–generous*). The second type contrasted traits that had the same evaluation but differed descriptively (e.g., *stingy–extravagant; thrifty–generous*). As expected, Felipe found that almost all inferences exhibited evaluative similarity for the first type of scales and descriptive similarity for the second type. Thus, like Peabody, he concluded that trait inferences are likely to follow descriptive similarity when possible; that is, descriptive similarity will be exhibited when the

13. Peabody also performed a factor analysis on the interscale correlations obtained from the 90×40 matrix of average trait inferences. He concluded that none of the major factors could be clearly identified as evaluative. Rosenberg and Olshan (1970), however, pointed out that this factor analysis did not provide a crucial test of Peabody's hypothesis (even if an evaluative factor had been found) since it was based primarily on unselected judgments. Moreover, Peabody (1970) agreed with Rosenberg and Olshan's argument that a factor analysis of these intercorrelations was inappropriate for his data. Using multidimensional scaling on similarity indices among Peabody's traits, Rosenberg and Olshan found a clear evaluative dimension. Similarly, A. L. Edwards (1969) also found an evaluative dimension when responses to an adjective checklist based on all of Peabody's traits were factor analyzed. However, as indicated above, the emergence of evaluative factors in these analyses do not contradict the notion that inferences are based on descriptive similarity.

traits share some degree of denotative or descriptive meaning. In the absence of descriptive similarity, inferences are likely to be based on evaluative similarity.

To summarize briefly, our discussion of research on trait inferences has revealed some important implications for an understanding of inferential belief formation. Much of the research in this area has attempted to provide systematic descriptions of the perceived relationships among traits. The most prominent finding has been that inferences from one trait to another follow a consistent pattern. Asch's findings concerning the effect on trait inferences produced by using *warm* or *cold* among a set of stimulus traits can be understood in terms of the dimensions underlying stimulus and response traits (Rosenberg, Nelson, and Vivekananthan, 1968). This stable factor structure was often interpreted as representing perceived personality dimensions. It was recognized, however, that the emerging factor structure could be interpreted as nothing more than instances of Osgood's general dimensions of meaning, and it became important to know whether the perceived trait relations (that produced the factor structure) were the result of evaluative or descriptive similarity. This issue in part led to the more general question raised earlier, namely, the extent to which inferential beliefs are formed on the basis of evaluative or probabilistic consistency. In contrast to the assumption of evaluative consistency that underlies much of attitude measurement as well as the notions of balance, congruity, affective-cognitive consistency, etc., research on trait inferences suggests that probabilistic rather than evaluative consistency plays a predominant role in inferential belief formation. Indeed, only in the absence of probabilistic consistency does evaluative consistency appear to determine the formation of inferential beliefs—at least so far as trait inferences are concerned.

Studies of interpersonal attraction. To further illustrate this point, it may be instructive to examine some experiments on interpersonal attraction which, although not directly concerned with belief formation, have obtained measures of belief in addition to measures of attitude. Generally speaking, subjects in these studies are given information about another person that is expected to influence their attraction to that person. On the assumption that individuals operate on the basis of evaluative consistency, a manipulation which is found to increase attraction toward the other person should also increase the belief that he has favorable attributes. This prediction follows from the assumption of an evaluatively consistent link between attitude and beliefs. In contrast, the assumption of probabilistic consistency implies that although certain information about another person may influence attraction, it will affect some but not all beliefs about that person. Specifically, its major impact should be on beliefs that are probabilistically related to the information provided.

Byrne (1969, 1971) and his associates have studied the relationship between similarity and interpersonal attraction. Subjects are given descriptions of a stranger's opinions or personality characteristics that differ in the extent to which they are similar to the subject's own opinions or personality traits. Subjects are

then asked to rate the stranger on Byrne's interpersonal judgment scale, which consists of two items designed to measure attraction and four "filler items." The latter items are measures of the subject's beliefs about the stranger's *intelligence, knowledge of current events, morality,* and *adjustment.* In a limited number of studies data concerning these filler items are reported, and the findings suggest a probabilistic inference process (see Fishbein and Ajzen, 1972). That is, inferences were made consistently only when the information provided about the stranger was descriptively similar to the judgment required. For example, beliefs about the other's knowledge of current events were affected only when the information dealt with political issues (Byrne, Bond, and Diamond, 1969) or a variety of opinions about social issues (Byrne and Ervin, 1969; Mascaro and Lopez, 1970). Beliefs about the other's adjustment were affected by similarity only in a study that provided information about the stranger's responses to a repression-sensitization test (Byrne and Griffitt, 1969). Finally, judgments of morality were affected by information about a candidate's stand on six social issues (Byrne, Bond, and Diamond, 1969).

Similar conclusions emerge from other experiments dealing with interpersonal attraction. For example, in two studies subjects provided information about themselves. A confederate then gave positive or negative personality evaluations of the subject, ostensibly on the basis of the information provided by the subject. The confederate who made positive evaluations was rated as higher in social sensitivity (Lowe and Goldstein, 1970) and as more intelligent and doing a better job (Sigall and Aronson, 1969). In contrast, when the subject merely overheard the confederate saying that he did or did not like the subject (Landy and Aronson, 1968), the confederate who expressed liking received higher ratings of kindness and friendliness, but no differential inferences were made with respect to eight other traits, including intelligence and sensitivity. Thus there is again little evidence that the subject's responses follow evaluative consistency. The generalizability of this conclusion to other processes of inferential belief formation will be examined below.

Cue Utilization

The basic research paradigm in studies of trait inferences is as follows: A stimulus person is described as possessing Trait A, and the subject is asked to indicate how likely it is that the stimulus person has Trait B. For example, the stimulus person might be described as *rich,* and the subject could be asked to indicate his subjective probability that the person is *intelligent.* An alternative strategy would be to provide information that the stimulus person earns a specified amount of money ($10,000, $15,000, or $20,000), and the subject could be asked to estimate how intelligent the person is (in terms of I.Q. scores). In this case, both the information provided and the subject's responses represent positions on different content dimensions. The subject's task is to infer the position of a stimulus person on one content dimension on the basis of information about his position on some other

content dimension. It seems reasonable to assume that this inference will be determined by the perceived relation between the two content dimensions. One possible way of estimating this relationship is to compute the correlation between the amount of money earned by the stimulus person and the subject's judgment of his I.Q.

This procedure can be extended by providing the subject with more than one item of information or *cue* about the stimulus person. In fact, in everyday life, people are often confronted with situations in which they must make some prediction or inference on the basis of several items of information or cues. For example, in hiring a job applicant, a personnel director may have to predict the applicant's probability of success on the job on the basis of information about the applicant's intelligence, previous experience, letters of reference, and amount of education. Similarly, stockbrokers make predictions about stock market behavior on the basis of such information as volume of trade, Dow Jones averages, etc.; doctors diagnose the probability that a tumor is malignant on the basis of X-rays, blood samples, size of tumor, etc. Indeed, most human judgments involve utilization of various items of information presumed relevant to the judgment in question.

When more than one cue is provided, it becomes possible to study a number of factors that may affect the inference process. For example, it is possible to examine the ways in which different cues are combined to arrive at the criterion judgment, to assess the effects of redundancy and inconsistency among cues, and to consider the extent to which cues are used efficiently to arrive at accurate inferences.

Multiple Regression Approach

For the purpose of studying such inference processes, a typical experiment on cue utilization might ask a subject to predict (infer) the intelligence quotients (I.Q. scores) for each of a group of college students on the basis of information about their grade point averages, aptitude test scores, credit hours attempted, and number of hours studied per week. The subject receives several quantified pieces of information (i.e., cues) about each stimulus person and is asked to infer that person's I.Q. score. As in natural situations, the cues take on different values for each college student.[14] Table 5.4 illustrates hypothetical profiles for 10 college students whose intelligence scores are to be inferred. Typically, a large number of such profiles are to be judged by each subject.

Once the subject has made his inferences, correlations can be computed between each cue dimension and the subject's predictions of intelligence. These correlations provide an indication of the extent to which a given cue dimension affects the inferential judgment. One major issue in research on cue utilization concerns

14. The cue values can be selected such that different cues are either independent, redundant, or inconsistent (i.e., uncorrelated, positively correlated, or negatively correlated) with each other.

Table 5.4 Hypothetical Cue Profiles for 10 College Students

Hypothetical student	Cues: Grade point average	Aptitude test score	Credit hours attempted	Hours studied per week
1	3.2	465	9	17
2	1.3	379	9	5
3	4.0	631	12	21
4	3.9	375	10	12
5	2.5	440	9	12
6	2.7	391	12	13
7	1.9	412	12	8
8	3.2	676	9	35
9	4.0	820	14	10
10	2.6	429	9	15

the process whereby different cues are combined to arrive at the inference. The simplest process is a linear combination of cues in which cue values are combined in an additive fashion. It is usually assumed that cue dimensions are given different weights, depending on their relevance for (or relation to) the criterion dimension. Much research on cue utilization has attempted to test the validity of this weighted linear model. Multiple regression analysis has served as the primary tool for this test since it provides estimates of cue weights as well as an index of the predictive accuracy of the weighted linear model.

To state it simply, the purpose of a multiple regression analysis is to provide an index of the degree of correlation between a *set* of independent variables (in this case the cues) and a given dependent variable (in this case the inference or criterion). That is, a multiple correlation coefficient is obtained which expresses the degree to which simultaneous consideration of the cues permits prediction of the inference. More specifically, each cue is given a weight (its *regression weight*) which represents its independent contribution to the prediction of the inference. The multiple regression analysis identifies optimal regression weights, that is, weights which lead to the most accurate prediction.[15] Further, multiple regression analysis combines the weighted cues in an additive fashion to arrive at the prediction. It follows that the multiple correlation is an indication of maximal predictability under the assumption of a linear (additive) model.

The correlations between cues and inference are used to compute regression weights. When the cues themselves are uncorrelated, the regression weight of each cue is identical to the cue's correlation with the inference. When the cues are correlated, these correlations are taken into account such that a given cue's regression

15. The optimization procedure is based on the least-squares criterion; that is, it minimizes the squared differences between predicted and obtained scores.

weight expresses the contribution this cue makes toward the prediction independent of all other cues. These regression weights are often assumed to indicate the relative weight placed on a given cue by the individual in making his inferences. Thus regression weights have been taken as indices of the relative *importance* of cues for a given inference.

Slovic and Lichtenstein (1971) have reviewed a large number of studies in which subjects were asked to use cues to make judgments about such things as personality characteristics, performance in college or on the job, physical and mental pathology, and legal matters. These studies demonstrate that inferential beliefs are predicted with considerable accuracy on the basis of a weighted linear combination of cues. Indeed, Slovic and Lichtenstein (1971, p. 678) concluded, "In all of these situations the linear model has done a fairly good job of predicting the judgments, as indicated by [multiple correlations] in the 80's and 90's for the artificial tasks and in the 70's for the more complex real-world situations."

Prediction versus description of the inference process. Despite the predictive success of the linear model, one cannot necessarily conclude that respondents actually combine cues in a linear fashion when forming inferential beliefs. In fact, there is some evidence that information may be combined in nonlinear or configural ways (e.g., Wiggins and Hoffman, 1968; Tversky, 1969; Einhorn, 1970).[16] However, research on configurality has repeatedly shown that even when there is evidence for configural processes, the linear model accounts for most of the variance in inferential beliefs (cf. Slovic and Lichtenstein, 1971; Goldberg, 1968). This suggests that accurate prediction by the linear model may not always provide a valid description of the inference process.

Other evidence suggesting that the linear model may not, in fact, accurately describe the actual inference process comes from studies in which subjects are asked to estimate the relative weight they placed on each cue in making their judgments. For example, Summers, Taliaferro, and Fletcher (1970) had subjects infer the level of socioeconomic development of 175 nations on the basis of four cues (amount of foreign investment in the country, amount of government influence over private enterprise, foreign aid received from United States, and number of socialist deputies or representatives in the nation's congress). After making these 175 judgments, subjects indicated the subjective weight they thought they placed on each cue by distributing 100 points among the four cues. The subjective weights were then compared with regression weights obtained in a multiple regression analysis. Large discrepancies were found between subjective and obtained weights. Moreover, although the multiple regression analysis indicated that approximately 75 percent of all subjects utilized three cues or fewer in making their

16. Wiggins and Hoffman (1968) found that in comparison with a simple linear model, a configural model improved predictions for many respondents. However, the increment in predictability was minimal.

judgments, almost every subject reported using all four cues. Thus the linear regression model and subjective reports suggest different combinations of cues in making the inferences. Such discrepancies between subjective and obtained weights have been found in many other studies (e.g., Hoffman, 1960; Hoepfl and Huber, 1970; Oskamp, 1962; Pollack, 1964). In comparison with the weights provided by the linear regression model, respondents tend to overestimate the importance they place on minor cues (i.e., on cues with low regression weights), and they underestimate their reliance on a few major cue dimensions.

Interestingly, there is some evidence (Slovic, Fleissner, and Bauman, 1972) that the correspondence between subjective and obtained weights decreases as the respondent becomes more experienced. Slovic and Lichtenstein (1971) have suggested that judgments become more automatic with experience and that experienced judges, therefore, may be less able to report the ways in which they utilize cues. This account implies that obtained weights represent the respondent's actual weights more accurately than his own subjective weights, but other investigators have argued that discrepancies between obtained and subjective weights reflect inadequacies of the linear model. Although the predictive accuracy of the linear model is not questioned, there is considerable doubt concerning its ability to provide an accurate *description* of the inference process. Indeed, there is a growing concern among investigators that the linear model may not provide an accurate descriptive account of inferential belief formation, despite its predictive success. As Slovic and Lichtenstein (1971, p. 683) have pointed out, "notions about nonlinear processes are likely to play an increasing role in our understanding of judgment despite their limited ability to outpredict linear models."

Whatever the descriptive utility of the linear model, the multiple regression approach to cue utilization has provided information concerning a number of important issues. One major focus of research in this area has been the systematic investigation of individual differences. There seems to be little doubt that respondents vary considerably in terms of the weights they place on given cues. That is, respondents tend to use the same information in very different ways, and thus the same information may lead to the formation of different inferential beliefs. Moreover, Wiggins, Hoffman, and Taber (1969) found that the weights placed on different cues were related to various personological characteristics of the judges. Subjects in this study inferred the intelligence of stimulus persons on the basis of nine cues, such as high school ratings, mother's education, emotional anxiety, and study habits. A factor-analytic procedure was used to identify types of judges who had made similar inferences; eight different groups or types of judges were thus isolated. Multiple regression analyses were performed for each group of judges by correlating the nine cue values associated with a given stimulus person with the group's mean judgment of the stimulus person's intelligence. As Table 5.5 shows, the different types of judges placed different weights on the nine cues, and the number of significant cues varied from one (Judge 7) to eight (Judge 6). Thus, for example, judges of Type 1 appeared to consider only high school rating and

Table 5.5 Multiple Correlations between Nine Input Cues and Mean Intelligence Judgments of 75 Profiles for Eight Discrete Subject Groups (From Wiggins, Hoffman, and Taber, 1969)

Profile cues	Zero-order correlations by groups							
	I ($n=45$)	II ($n=28$)	III ($n=53$)	IV ($n=10$)	V ($n=5$)	VI ($n=2$)	VII ($n=1$)	VIII ($n=1$)
1. High school rating	.97‡	.34†	.45‡	.61‡	.38‡	.43‡	.05	.70‡
2. Status	.08	.08	.20	.05	.81‡	.35†	.01	.10
3. Self-support	−.03	−.09	.07	.08	.03	.25*	.10	.17
4. English effectiveness	.16	.90‡	.42‡	.24*	.32†	.27*	.02	.07
5. Responsibility	−.10	.04	.46‡	−.01	.20	.18	.01	−.32†
6. Mother's education	.05	.07	.15	.26*	.12	.39‡	−.06	−.06
7. Study habits	.00	.03	.46‡	−.08	−.09	.24*	.00	.08
8. Emotional anxiety	.08	.09	.11	.05	.11	.29*	.99‡	.07
9. Credit hours	.23*	.27*	.24*	.75‡	.17	.44‡	−.09	.04
Multiple correlation	.99‡	.99‡	.98‡	.95‡	.97‡	.93‡	.99‡	.80‡

* $p < .05$.
† $p < .01$.
‡ $p < .001$.

number of credit hours attempted, whereas judges of Type 5 placed primary emphasis on status, high school rating, and English effectiveness.[17]

We can gain some understanding of the basis for these differences by examining certain personological characteristics of the different types of judges. Types 1 and 2, who were found to place most weight on seemingly relevant cues (high school rating, English effectiveness, credit hours attempted), tended to be intelligent and low in ethnocentrism. In contrast, judges of Type 5, who placed heavy emphasis on the "character" variables of responsibility and study habits (cues which are seemingly irrelevant to the judgment of intelligence), were high on authoritarianism and religious conventionalism. These results support our earlier argument that the formation of inferential beliefs involves factors other than those in the immediate stimulus situation.

In addition to showing that different individuals place different weights on the various cues, research using the multiple regression approach has also made it clear that the weights of cues may vary with the judgment being made. Thus, for example, if Wiggins, Hoffman, and Taber had asked their subjects to judge each stimulus person's grade point average or motivation (rather than his intelligence),

17. One additional aspect of these results is worth noting. All nine cues as well as the criterion can be placed along an evaluative dimension. Nevertheless, there is little evidence for the notion of evaluative consistency since evaluative consistency implies that all cues should contribute to the inferences.

the weights of the cues would have shifted considerably for each type of judgment. Such effects are to be expected since, as we saw in our discussion of trait inferences, a given stimulus trait (or cue) tends to exhibit different degrees of relationship with different response traits (or criteria). This supports our argument in Chapter 4 that the effect of a given independent variable on a measure of belief cannot be generalized to measures of other beliefs or to measures of attitudes or intentions.

Analysis of Variance Approach

A second major approach to the investigation of the ways in which cues are combined or integrated has relied on the analysis of variance (ANOVA). For example, N. H. Anderson (1970, 1971a) has relied on the ANOVA model in developing his theory of information integration. This theory is concerned with two interrelated problems. First, Anderson has used the ANOVA model to obtain scale values of the various cues as well as to assign appropriate weights. Second, integration theory uses the ANOVA approach to test certain models of information integration.

Basically, an analysis of variance partitions the total variance in the dependent variable (the judgments) and provides estimates of the amount of this variance due to each of the stimulus dimensions or factors (main effects) and their interactions. Further, the analysis of variance tests the statistical significance of these main effects and interactions. If a simple additive or linear combination of cue values determined judgments (as is suggested by a linear regression model), only the main effects should be significant. Findings of significant interactions constitute evidence for configurality in judgments.

In the multiple regression approach to cue utilization discussed above, investigators have typically selected cue profiles that were descriptive of real persons or objects. Indeed, they have frequently employed available cue profiles. Thus many studies employing the multiple regression approach have used the personality profiles of 861 psychiatric patients analyzed by Meehl (1959). In contrast, when the analysis of variance (ANOVA) paradigm is used, investigators must construct all possible combinations of cue values. Use of such an orthogonal factorial ANOVA design ensures that the cue dimensions are unrelated since every cue value is paired with every other cue value.

Most research on Anderson's integration theory has focused on attitudes rather than beliefs, and this research will be reviewed in Chapter 6 on attitude formation. In one of the few studies applying integration theory to inferential beliefs, Himmelfarb and Senn (1969) had subjects infer the social class of stimulus persons on the basis of three cues: income, education, and occupation. Using four values for each cue, they constructed the 64 possible cue profiles shown in Table 5.6. For example, the cell in the upper left-hand corner represents a stimulus person who is a tobacco laborer with fourth-grade education and a yearly income of $2,500. The cell in the bottom row of the third column represents a jeweler with a professional degree and a yearly income of $20,000. Note that some of the profiles in

Table 5.6 ANOVA Design Used by Himmelfarb and Senn (1969)

Education	Yearly income	Occupation: Tobacco laborer	Taxicab driver	Jeweler	Banker
Fourth grade	$ 2,500				
	4,500				
	10,000				
	20,000				
Ninth grade	2,500				
	4,500				
	10,000				
	20,000				
College degree	2,500				
	4,500				
	10,000				
	20,000				
Professional degree	2,500				
	4,500				
	10,000				
	20,000				

Table 5.6 are highly unlikely and would usually not have been included in a multiple regression analysis.[18] Among such unlikely stimuli are a "banker with a professional degree making $2,500 a year" and a "tobacco laborer with a fourth-grade education and a yearly income of $20,000."

The 70 subjects in Himmelfarb and Senn's experiment judged each of the 64 stimulus profiles twice. Separate analyses of variance performed for each of the 70 subjects revealed significant interactions for 22 subjects, indicating that the simple linear model was inadequate for at least some respondents. Himmelfarb and Senn also computed an analysis of variance over all subjects and found that all main effects as well as the pooled interaction[19] were significant. Thus each

18. Indeed, Brunswik (1952) has argued that judgments should be studied in realistic settings, and the inclusion of unrealistic profiles may be disruptive of the very process they were meant to disclose.

19. The pooled interaction is the amount of variance associated with all four interaction terms in the analysis (i.e., occupation × income, occupation × education, income × education, and occupation × education × income).

dimension made a significant contribution to judgments of social status, but prediction could be improved by taking interactions into account. This finding again indicates that a simple linear model is not an accurate description of the way in which some of the subjects utilized information to make judgments about social status.

Himmelfarb and Senn attempted to explain the deviations from linearity. In accordance with Anderson's integration theory, they used row and column means as estimates of each cue's scale value. For example, the mean judgment of social status for all stimulus persons who are tobacco laborers (i.e., all stimuli in the first column of Table 5.6) is used as an estimate of the scale value of the cue "tobacco laborer." Similarly, the mean judgment of social status for all persons earning $10,000 is used as an estimate of the value placed on the cue "$10,000." Assuming equal weights, they combined these scale values in a linear fashion to compute predicted judgments for each stimulus person. By examining those profiles that exhibited the greatest discrepancies between predicted and obtained judgments, Himmelfarb and Senn discovered a "contrast effect." When one cue value was high and the other two low (or when one was low and two high), the obtained judgment was higher (or lower) than the judgment predicted by a linear model with equal weights.[20] Thus the judgments indicated that more weight was placed on the inconsistent cue.

There is clearly some evidence for nonlinear combinations of cues in the analysis of variance paradigm, but here, as in studies using multiple regression analyses, much of the variance can be accounted for by a linear combination rule.

In conclusion, research on cue utilization suggests that a person's inferences are derived in a consistent and predictable manner from the information available to him. When either a multiple regression or an analysis of variance approach is used, integration of diverse items of information is found to be consistent with a weighted linear model. That is, by making the assumption that each cue dimension is given a weight and that the weighted cue values are combined in an additive fashion, one can make highly accurate predictions of the individual's inferences. However, there is considerable doubt whether the weighted linear model is an accurate description of the actual inference process. Empirical evidence suggests that people often combine information in nonlinear ways to arrive at their judgments, in spite of the fact that a weighted linear model accounts for much of the variance in these judgments. A more detailed comparison of linear and non-

20. This contrast effect contradicts the findings of an earlier study by Anderson and Jacobson (1965), in which the inconsistent cue was found to be "discounted"; that is, it was given less (not more) weight than expected on the basis of a linear model. However, in contrast to the judgments of belief content required of subjects in the Himmelfarb and Senn experiment, subjects in the earlier study made evaluative judgments. Thus, again, a given finding with respect to attitudinal responses may not be generalizable to a belief dimension. Moreover, in a second study reported by Himmelfarb and Senn (1969) neither a contrast nor a discounting effect was found.

linear models will be found in our discussion of attitude formation in Chapter 6.

Irrespective of the descriptive validity of a linear model, research on cue utilization, like research on trait relations, suggests that an individual's inferences are based on fairly stable and internally consistent patterns of relations among beliefs. Since studies on cue utilization have often obtained judgments about objects other than persons, subjects may be viewed as holding not only implicit theories of personality but also implicit theories about the attributes of objects, institutions, or events. By focusing on individual differences, cue utilization research has demonstrated that the perceived relations among attribute dimensions vary across individuals. This raises the question as to how a person learns the relations between beliefs.

Learning to Make Accurate Inferences

Thus far, we have been concerned primarily with descriptive accounts of inferential belief formation, and we have not paid much attention to the accuracy of the inferences. Festinger (1954) has suggested that people strive to hold correct opinions or beliefs about the world. Indeed he "posited the existence of a drive to determine whether or not one's opinions were 'correct' " (p. 118). He further argued that when objective, nonsocial means are not available, people test the accuracy of their opinions by comparing them with the opinions of others. It follows that individuals will tend to revise their beliefs as a function of the positive and negative feedback they receive from the social or nonsocial environment. For example, a person may believe that a certain politician favors increased military spending. He can attempt to assess the accuracy of this belief by discussing the candidate's position with his friends, coworkers, etc., who will agree or disagree with him. Depending on the nature of this feedback, he may revise his belief. Alternatively, he can examine the politician's voting record or attend a lecture given by the politician, thus receiving direct information concerning the degree to which his belief is correct. Festinger (1954) has argued that a person will compare his beliefs with those of others primarily when objective criteria are not available. As a result of this feedback and revision process, a person will tend to hold beliefs that reflect physical realities or, when objective physical criteria are unavailable, beliefs that tend to agree with those of certain relevant others.[21]

Two lines of research have been concerned with the effects of feedback on inferential beliefs: multiple cue learning and concept formation.

Multiple Cue Learning

We have seen that multiple regression analyses have been used to estimate the relative weights that an individual places on different cues in arriving at an inference. These weights were found to permit accurate prediction of the individual's

21. Among other things, Festinger's theory of social comparison processes attempts to identify the relevant others who are used for comparison purposes.

inferences on the basis of a weighted linear model. We can now ask whether it is possible to provide the individual with information that will change these weights such that he will make more accurate judgments of the criterion.[22]

In multiple cue learning, a subject is shown a set of cues describing some object and is asked to make an inferential judgment about the object. After each judgment he is given feedback as to the accuracy of his judgments. Sometimes subjects are simply told whether their inferences were correct or incorrect. Alternatively, they may be given information about the degree to which they have placed appropriate weights on the various cues (Hammond and Summers, 1972).

Brunswik's lens model. Clearly, multiple cue learning requires that some "objective" criteria or "distal variables," such as the stimulus person's actual I.Q. score or grade point average, be available. The correlation between a given cue dimension and this objective criterion is taken as an index of the cue's relevance with respect to the criterion; it has been called the cue's *ecological validity*. In addition, it is possible to compute the correlation between the individual's prediction and the actual criterion. This correlation, called the *achievement index,* reflects the accuracy of the individual's judgment. The relations between cue values, judgments, and criterion scores are part of a more complex set of relations defined by Brunswik's lens model (cf. Brunswik, 1955, 1956; Dudycha and Naylor, 1966). Figure 5.4 illustrates Hammond, Hursch, and Todd's (1964) conceptualization of

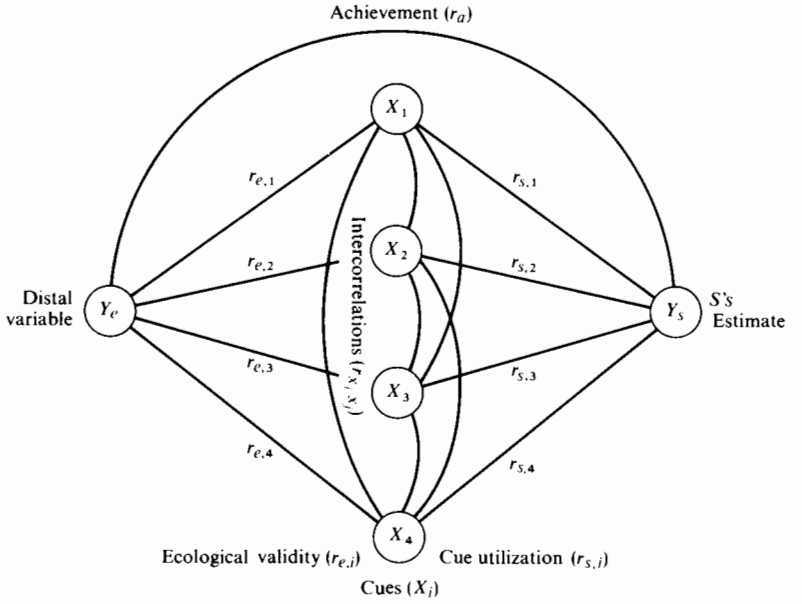

Fig. 5.4 The Lens Model. (From Hammond, Hirsch, and Todd, 1964.)

22. Such changes may occur, whether or not the person can accurately report the regression weights.

Brunswik's model. The model deals with relations among cues, with relations of cues to the subject's estimates (cue utilization) and to the criterion (ecological validity), and with the relation between estimates and criterion (achievement).

Early studies examined the degree to which subjects can learn to form more accurate inferential beliefs. Several studies have shown that accuracy increases as the subject's weights (i.e., his cue utilizations) come to approximate the ecological validity of cues. When subjects are given feedback as to the accuracy of their judgments, they progressively learn to place more appropriate weights on the various cues (e.g., Azuma and Cronbach, 1966; Lee and Tucker, 1962; Summers, 1962). With respect to the type of feedback given, Hammond and Summers (1972) found that telling subjects whether their inferences were correct or incorrect is not so effective as giving them information about the degree to which they placed appropriate weights on the various cues.[23]

Hammond and Summers (1972) have also shown that accuracy is influenced not only by appropriate weighting of cues but also by the degree to which these weights are used in a systematic fashion. Even after a person has learned to assign appropriate weights to the cues, such random factors as boredom, mood, and inattention can result in inconsistent application of cue weights, thus adding error to the judgment and reducing accuracy. Consistent with this notion, Goldberg (1970) showed that the multiple regression equation (which uses the subject's regression weights in a systematic fashion) enabled predictions of the criterion which were more accurate than the subject's own predictions. Hence the "model of man" (i.e., use of a linear regression model) tends to be more consistent and accurate than man himself.[24]

According to Hammond and Summers (1972), a distinction must be made between *obtaining* "knowledge" or information and *using* that knowledge appropriately. A similar distinction has been made in research on concept formation between learning the *attributes* of a concept (i.e., learning the relevant cues) and learning the *rule for combining* these attributes.

Concept Formation

The basic research paradigm in studies on concept formation is very similar to multiple cue learning. The subject is shown a multidimensional stimulus, such as

---

23. Studies of multiple cue learning have manipulated various aspects of ecological validity to examine effects on accuracy. For example, the relevance of cues, the nature of the cue-criterion relation, redundancy and inconsistency among cues, etc., have been found to influence accuracy of inferences and the number of trials or time required to achieve a given level of accuracy. For a review of this literature, see Slovic and Lichtenstein (1971) and D. M. Johnson (1972).

24. This finding led Goldberg to suggest a "bootstrapping" strategy to systematize prediction by using expert judges to obtain regression weights and then use the regression equation to make predictions. This procedure is similar to an approach taken by W. Edwards (1962) and others in attempts to systematize use of information in Bayesian decision making, to be discussed below.

a card containing three large blue circles. The dimensions along which stimuli vary in studies of concept formation (e.g., color, size, shape, numerosity) are comparable to cues (such as grade point average, number of credit hours attempted, etc.) in studies of multiple cue learning. The subject is then asked to associate a concept label with the stimulus or to indicate whether the stimulus is or is not an instance of a particular concept class. For example, the investigator might define the concept YAF as "three blue objects." Any stimulus displaying three blue objects—whether large or small, circles or squares—is considered a member of the concept class (is a positive instance of YAF); any stimulus that fails to display the two attributes "three" *and* "blue" is a negative instance. The investigator may define his concept in terms of any number of attributes under a variety of combination rules. In our example, the combination rule was *conjunctive* (i.e., three *and* blue). The disjunction "three" *or* "blue" would be an alternative combination rule defining a different concept. In this case, any stimulus displaying either all blue objects or any three objects would be a positive instance of the concept class. Following each judgment, the subject is told whether his judgment was correct or incorrect.

This research paradigm allows systematic investigation of the ways in which a person acquires a concept in many real-life situations. For example, a child is exposed to various moving objects, and he is told that some of these moving objects are positive instances of the concept BUS and that others are negative instances. He somehow uses this information to identify the defining attributes of BUS.

The subject's task in a concept formation study, then, is to discover the attributes and the combination rule that define a given concept. In this framework, belief formation can be viewed as a problem-solving process. Individuals use information to revise their beliefs or hypotheses concerning relevant attributes and combination rules. We can now see that multiple cue learning is also a problem-solving situation in which subjects revise their hypotheses about the relevance of cue dimensions for the criterion (i.e., about each cue's ecological validity).

As in multiple cue learning, feedback leads to successive revisions in hypotheses until the concept is correctly identified. For example, a subject might be told that the concept is defined by a conjunctive rule and that the stimulus complex "two large green squares" is a positive instance of the concept in question. At this stage, he knows that all other attributes (e.g., one, small, red, or circle) are not part of the concept. He may at this stage form the hypothesis that the concept is "square and green." If he now learns that the stimulus "two large *red* squares" is also a positive instance of the concept, he will tend to eliminate color as a relevant attribute. On the other hand, if he learned that "two large red squares" was a negative instance, he would know that "green" *is* a relevant attribute. Investigations using positive and negative instances of this kind have shown that both types of information permit subjects to form the appropriate concept. Depending on the combination rule, positive instances may be more or less informative than negative instances. Specifically, positive instances tend to be

superior when the concept involves a conjunctive rule (e.g., red *and* square) whereas negative instances are superior with disjunctive rules (e.g., red *or* square). In either case, however, subjects can learn to use both types of instances with equal effectiveness (see Hovland and Weiss, 1953; Freibergs and Tulving, 1961; Chlebek and Dominowski, 1970).

Not surprisingly, studies of concept formation have manipulated many of the same stimulus variables considered in studies of multiple cue learning. Thus stimulus attributes have been manipulated in terms of redundancy, consistency, number of relevant and irrelevant attributes, the rules relating attributes to concepts, etc. To a large extent, emphasis has been placed on the effects of these factors on speed of concept acquisition (e.g., number of trials required to identify the concept) or number of errors. (For reviews of this literature, see D. M. Johnson, 1972, and Bourne, Ekstrand, and Dominowski, 1971.) Although speed of learning and number of errors are relevant in a problem-solving context, these measures provide little direct information concerning processes underlying formation of inferential beliefs. These measures deal with the accuracy of beliefs rather than with the ways in which inferences are made, whether the inferences are accurate or not.

Some aspects of research on concept formation, however, are more directly relevant for an understanding of inferential belief formation. Concept formation is often viewed as involving a process of logical reasoning whereby alternative hypotheses are successively eliminated (cf. Levine, 1966; Huttenlocker, 1967; Bourne, 1963). Much of the concept formation research has attempted to identify and describe the kinds of strategies that subjects use in their attempts to discover the "correct" concept [25] (e.g., Bruner, Goodnow, and Austin, 1956; Levine, 1966; Laughlin and Jordan, 1967; Laughlin, 1965, 1966; Bourne, 1963). These strategies or logical reasoning processes employed by persons when they acquire a concept are examples of formal coding systems (Bruner, 1957) that may provide the basis for inferential belief formation.

Formal Inference Models

We have seen that inferences are often based on perceived relations among beliefs. These lawful relations among beliefs tend to be based on probabilistic consistency. Various models derived from probability theory deal explicitly with the relations between subjective probabilities (i.e., beliefs). Only recently, however, have models of this type attracted the attention of investigators in the attitude area. Prior to the development of mathematical probability theory, philosophers concerned with logic and syllogistic reasoning also attempted to formalize the inference process, and we therefore turn first to a brief discussion of syllogistic reasoning.

---

25. Sometimes subjects are asked to provide verbal reports of their hypotheses at various stages of the concept formation process.

Syllogistic Reasoning

Logical syllogisms. In our discussion of concept formation, we emphasized the processes whereby a person solves the problem of identifying a given concept. One possible interpretation of the person's behavior in concept formation or problem-solving situations is that he attempts to eliminate alternative hypotheses by some process of logical reasoning. The approach to an understanding of such processes has been for the experimenter to specify a logical analysis of the necessary and sufficient steps or operations for solving the problem. "The logical sequence then serves as a normative model of the subject's behavior and the experimenter looks for evidence of this logic in the subject's actual behavior." (Bourne, Ekstrand, and Dominowski, 1971, p. 230)

The problem of drawing valid conclusions from certain premises has occupied philosophers for a long time, and their efforts have resulted in a set of formal rules for making such inferences. Research concerning the degree to which individuals draw valid conclusions as prescribed by these formal models has centered around the logical syllogism. A syllogism consists of three belief statements or propositions, two of which serve as premises and the third as a conclusion. For instance, consider the following three propositions.

Premise 1: Communists are atheists.

Premise 2: Russians are Communists.

Conclusion: Russians are atheists.

Belief statements of this kind link an object (O) to an attribute (X). Within the syllogistic framework, four forms of belief statements have been considered.

1. Universal affirmative: All O is X.
2. Universal negative: No O is X.
3. Particular affirmative: Some O is X.
4. Particular negative: Some O is not X.

One standard syllogistic figure involving the conclusion O is X consists of a minor premise linking O to some middle term M, a major premise linking M to X, and the conclusion O is X. Indeed, the example above concerning Russians and atheism can be stated symbolically as follows:

Major premise: M is X (Communists are atheists).

Minor premise: O is M (Russians are Communists).

Conclusion: O is X (Russians are atheists).

One difficulty with this example is its failure to specify whether its affirmative propositions are universal or particular, thus making impossible the determination of the conclusion's validity. The conclusion "All Russians are Communists" would be valid only if both major and minor premises were stated in universal affirmative form. Clearly, by varying the forms of belief statements (i.e., their

universality and affirmativeness), one can construct many different syllogisms, only some of which have valid conclusions.

Similarly, it is possible to construct different types of syllogisms by varying the position of the middle term in major and minor premises. Four basic configurations, or types, can be identified.

	Type 1	*Type 2*	*Type 3*	*Type 4*
Major premise:	*M* is *X*.	*X* is *M*.	*M* is *X*.	*X* is *M*.
Minor premise:	*O* is *M*.	*O* is *M*.	*M* is *O*.	*M* is *O*.
Conclusion:	*O* is *X*.	*O* is *X*.	*O* is *X*.	*O* is *X*.

Stated in universal affirmative form, only the first type of syllogism is valid. When forms other than the universal affirmative are used, the different types of syllogisms lead to some valid and some invalid conclusions. Further, it can be impossible to determine whether the conclusion is valid or invalid.

Syllogistic reasoning has often been investigated by varying the type of syllogism and the form of the propositions. A judgment is typically elicited in one of two ways. First, the complete syllogism is presented, and the subject is asked to indicate whether the conclusion is valid or invalid (true-false format). Alternatively, the two premises are presented together with the set of four forms of the conclusion, and the subject is asked to judge which form is valid (in a multiple-choice format); he is usually given the additional alternative "No valid conclusion is possible." Experimental research indicates that more invalid conclusions are reached with some types of syllogisms and proposition forms than with others (Woodworth and Sells, 1935; Morgan and Morton, 1944; Chapman and Chapman, 1959; Roberge, 1970). Further, with the multiple-choice format, different subjects tend to make similar systematic errors. For example, given a syllogism where no valid conclusion is possible, there is usually strong agreement among subjects that one of the incorrect alternatives is valid (Roberge, 1970). Another factor that has influenced judgments of conclusion validity is the content of the propositions. Syllogisms have been presented in terms of abstract symbols (e.g., "All *x* are *y*"), in terms of nonsense syllables (e.g., "All DAB are ZIG"), or in terms of meaningful statements (e.g., "All Russians are Communists"). Although the same type of errors tends to be found with the different formats, the introduction of meaningful propositions results in certain unique aspects. For example, the desirability or controversiality of the conclusion has been found to affect judgments of validity (e.g., Janis and Frick, 1943; Lefford, 1946).

Like studies of multiple cue learning and concept formation, studies of syllogistic reasoning have dealt with problem-solving situations, and they are only indirectly relevant for an understanding of the processes underlying inferential belief formation. However, by carefully examining the kinds of errors that subjects typically make, one can gain some information concerning their reasoning processes. Note that within a problem-solving situation, a person may agree that a given conclusion can be validly derived from the premises, even though he may

not *agree* with either the premises or the conclusion. Indeed, as noted above, some investigators have examined the degree to which a person's own beliefs or attitudes affect his judgments of validity. Although such effects have sometimes been found, there is evidence that subjects who score high on an ordinary test of reasoning ability (i.e., one using abstract symbols or noncontroversial material) are most able to discount their own biases when judging the validity of controversial conclusions (Feather, 1964b). This finding, like others noted above, suggests that subjects tend to make inferences along lines of probabilistic consistency and will make judgments along evaluative lines only when there is no basis for probabilistic consistency or when they fail to perceive such a basis.

This question concerning effects of beliefs and attitudes on reasoning emphasizes the distinction between a problem-solving approach and our interest in inferential belief formation. To study the formation of inferential beliefs, one would be concerned not with the effects of various factors on a subject's ability to judge the *validity* of a given conclusion, but rather with their effects on the degree to which he *agrees* with the conclusion. Thus one would be interested in knowing whether the type of syllogism or form of propositions influences acceptance of the conclusion. Similarly, one would investigate the extent to which acceptance of the conclusion is influenced by the degree to which a person agrees with (or believes) the major and minor premises or by his judgment of the conclusion's desirability.

Verbal generalizations. Abelson and his associates (Gilson and Abelson, 1965; Abelson and Kanouse, 1966; Kanouse, 1971) have investigated the likelihood that subjects will *accept* some general conclusion on the basis of limited information. "In these studies respondents were presented with simple assertions (i.e., propositions) containing a subject, verb, and object. In addition they were given mixed evidence bearing on the truth of the assertion; that is, some of the evidence supported the assertion while some contradicted it. The respondent's task was to indicate whether he felt that the assertion was true or false in the light of the given evidence." (Kanouse, 1971, p. 2) The evidence consisted of three propositions pertaining either to the subject or the object of the conclusion. Further, the information was selected such that subjects could either make an induction by inferring the conclusion from specific instances or make a deduction by inferring the conclusion from more general qualities. Examples of the four kinds of quasi-syllogisms used by Abelson and Kanouse (1966) are presented in Table 5.7. Respondents were given a large number of such items, with subjects, verbs, and objects systematically varied from item to item. "Instructions emphasized that the task should not be regarded as a test of logic but that the subjects' answers should be based on their intuitive reactions to the evidence." (Kanouse, 1971; p. 2) Nevertheless, it appears that a certain degree of problem solving is involved in this task; that is, subjects are asked to judge whether the conclusion appears reasonable in light of the evidence presented.

Results of several studies indicated that these judgments were influenced primarily by three factors: whether the item involved induction or deduction,

Table 5.7 Quasi-Syllogisms (from Abelson and Kanouse, 1966)

Form I.	Altogether there are three kinds of tribes: Southern, Northern, Central
	Southern tribes do not have sports magazines
	Northern tribes have sports magazines
	Central tribes do not have sports magazines
	Do tribes have sports magazines?
Form II.	Altogether there are three kinds of magazines: Sports, News, Fashions
	Southern tribes do not have sports magazines
	Southern tribes do not have news magazines
	Southern tribes have fashion magazines
	Do Southern tribes have magazines?
Form III.	All candidates are ambitious, long-winded, outgoing individuals
	Ambitious individuals produce choral music
	Long-winded individuals produce choral music
	Outgoing individuals do not produce choral music
	Do candidates produce choral music?
Form IV.	All choral music is agreeable, rhythmic, noninstrumental sound
	Candidates produce agreeable sound
	Candidates do not produce rhythmic sound
	Candidates produce noninstrumental sound
	Do candidates produce choral music?

whether the evidence was subject-specific or object-specific, and the kind of verb used. For inductive items, object-specific evidence produced greater agreement than subject-specific evidence, whereas the opposite was true for deductive items. Further, the amount of agreement produced by inductive or deductive items depended on the nature of the verb in the conclusion. Attempts to explain the unanticipated strong effect of the verb have focused on the universality of propositions. We have shown earlier that statements such as "candidates produce choral music" are ambiguous in that they do not specify whether the proposition is universal or particular. Clearly, a given set of evidence may be sufficient for a person to agree that "*some* candidates produce choral music" but not that "*all* candidates produce choral music." Abelson and Kanouse (1966) found that different verbs influence the amount of evidence required to produce agreement with the ambiguous conclusion.[26]

26. Kanouse (1971) has suggested an alternative explanation, according to which different verbs lead respondents to assign different "implicit quantifiers" (i.e., all, some, a few, etc.) to subjects and objects.

Although these studies are suggestive of the ways in which relations among propositions can be investigated, and although they have produced some interesting results, the processes of probabilistic inferences have not been formalized. In our opinion, research concerning these questions could profit from an application of some more formal probabilistic models which provide precise statements of the relations between premises and conclusions. Such models make explicit the central question asked in the studies by Abelson and his associates: What is the probability of accepting a conclusion, given certain types of evidence? Let us thus consider some probabilistic formulations that have been applied in research on inferential belief formation.

Probability Models

Based on mathematical probability theory, various models have been employed to account for relationships between subjective probabilities.[27] These models are *normative* rather than *descriptive;* that is to say, they prescribe what relations should exist between probabilities or how beliefs ought to change in light of new information. Thus probability models describe the belief structures of "rational" or "logically consistent" persons rather than those of actual persons. Nevertheless, it is possible to use probability models as first approximations for a psychological theory of inferential belief formation (Peterson and Beach, 1967). Probabilistic consistency, therefore, refers to the degree to which a person's inferences correspond to the relationships specified by probability theory; deviations from the normative model indicate that such "nonrational" factors as the person's attitudes or personality characteristics may influence his probabilistic judgments or beliefs.

Several investigators have attempted to test the general hypothesis that relations among subjective probabilities obey objective-probability laws. For example, according to probability theory the conjunctive probability that both A and B are true is represented by Eq. 5.1,

$$p(A \cap B) = p(A|B)p(B), \tag{5.1}$$

where $p(A \cap B)$ is the conjunctive probability of A and B; $p(A|B)$ is the conditional probability of A, given that B is true; and $p(B)$ is the probability of B.

The conjunction $p(A \cap B)$ can also be written

$$p(A \cap B) = p(B|A)p(A). \tag{5.2}$$

Equations 5.1 and 5.2 imply that

$$p(A|B)p(B) = p(B|A)p(A). \tag{5.3}$$

27. Following Savage (1954) it has usually been assumed that subjective probabilities have the same mathematical properties as the objective probabilities of mathematical probability theory. Thus, it is usually assumed that subjective probabilities range from 0 to 1 and that the subjective probabilities of a set of mutually exclusive and exhaustive events sum to 1.

According to probability theory, the disjunctive probability that either A or B is true is given by Eq. 5.4,

$$p(A \cup B) = p(A) + p(B) - p(A \cap B), \tag{5.4}$$

where $p(A \cup B)$ is the disjunctive probability of A or B.

Wyer (1970a) investigated the descriptive accuracy of Eqs. 5.2 and 5.4. He obtained measures of subjective probabilities for each term in these equations on a scale ranging from 0 (*extremely unlikely*) to 10 (*extremely likely*). The estimates were divided by 10 so that the subjective-probability measures could range from 0 to 1. Subjects read nine hypothetical situations, and their subjective-probability estimates were based on these descriptions. For example, one situation dealt with the reelection of a political candidate (Governor Smith). Two belief statements were of central importance.

A: Governor Smith will be reelected.

B: State aid to education will be increased.

Subjects were asked to estimate the following probabilities.

$p(A)$	The probability that Governor Smith will be reelected
$p(B)$	The probability that state aid to education will be increased
$p(B\|A)$	The probability that if Governer Smith is reelected, state aid to education will be increased
$p(B\|\overline{A})$	The probability that if Governor Smith is not reelected, state aid to education will be increased
$p(A \cap B)$	The probability both that Governor Smith will be reelected *and* that state aid to education will be increased
$p(A \cup B)$	The probability either that Governor Smith will be reelected *or* that state aid to education will be increased *or* that both events will occur

Wyer's results supported the notion that subjective probabilities follow the laws of objective probabilities. A correlation was computed between $p(A \cap B)$ and $p(B|A)p(A)$ in each of the hypothetical situations; the average correlation was .61. Further, there was considerable correspondence between $p(A \cup B)$ and $p(A) + p(B) - p(A \cap B)$, as suggested by Eq. 5.4. Similar results were reported by Wyer and Goldberg (1970).

The descriptive accuracy of Eq. 5.3 was examined by Peterson *et al.* (1965). These investigators looked at inferences among a set of 20 personality traits, such as dishonest, brave, witty, weak. For each trait, subjects provided estimates of unconditional or prior probabilities by indicating the number of people in 100 that could be described correctly by the trait. Further, estimates of conditional probabilities for all 190 possible trait pairs were also obtained by asking the following question: 100 persons are known to be *Trait A* (e.g., weak). How many of them would you expect to be *Trait B* (e.g., witty)? Thus, for each subject

95 estimates of $p(A|B)$ and 95 estimates of $p(B|A)$ were available. A correlation was computed for each subject across the 95 values of $p(A|B)p(B)$ and $p(B|A)p(A)$; the mean correlation for the total sample of 12 subjects was .67. In a second study, a correlation of .92 was reported for a different estimation task.

It thus appears that probability formulations of this kind provide fairly accurate descriptions of the relations among subjective probabilities. Reviewing the literature on intuitive statistical inferences, Peterson and Beach (1967) concluded that "probability theory and statistics can be used as the basis for psychological models that integrate and account for human performance in a wide range of inferential tasks. . . . Experiments that have compared human inferences with those of statistical man show that the normative model provides a good first approximation for a psychological theory of inference." (pp. 29 and 42)

Wyer (1970b) has shown that probability notions can be used to provide explicit quantitative definitions of several concepts discussed previously. Thus *redundancy* of two pieces of information A and B is defined by the degree to which $p(A \cap B)$ is greater than $p(A)p(B)$. Conversely, *inconsistency* of two pieces of information is defined by the degree to which $p(A \cap B)$ is smaller than $p(A)p(B)$. These definitions can be derived from Eq. 5.1. When A and B are independent pieces of information, that is, when they are neither redundant nor inconsistent,

$$p(A|B) = p(A). \tag{5.5}$$

Thus, if we assume independence, Eq. 5.1 can be rewritten

$$p(A \cap B) = p(A)p(B). \tag{5.6}$$

According to Wyer, when $p(A \cap B) > p(A)p(B)$, the two pieces of information are redundant since their joint probability exceeds the probability that would be expected if they were independent. Similarly, when $p(A \cap B) < p(A)p(B)$, the two pieces of information must be inconsistent. Wyer also used probability notions to define *novelty;* specifically, he defined novelty of two pieces of information A and B as the probability of their disjunction, that is, $p(A \cup B)$. Looking at Eq. 5.4, we can see that novelty of two events is thus defined as the sum of the information provided by each event in isolation minus the information common to the two events. The utility of these definitions will be discussed in later chapters.

A Model of Logical Consistency

One of the first applications of probability notions in the attitude area was McGuire's (1960a, b, c) attempt to provide a quantitative definition of consistency among beliefs by developing a model that combined formal logic and probability theory. The model was developed and applied in the context of logical syllogisms. McGuire used syllogisms of Type 1 and argued as follows: "Consider three issues so interrelated that a person taking Stands a and b on two of these issues would, if he were to be consistent, be logically required to take Stand c on the third. That is, the three stands are in a syllogistic relationship, with c following as a valid con-

clusion from the conjunction of a and b. If the person's opinions on these issues are obtained by having him indicate his adherence to each of the Stands, a, b, and c, on a probabilistic scale, we can specify that to be completely consistent, these probabilistically scaled opinions must be interrelated in the form

$$p(c) = p(a \cap b) + p(k)(1 - p(a \cap b)), \tag{5.7}$$

where $p(k)$ is his opinion of the probability of c on bases other than the conjunction of a and b." (McGuire, 1960b, p. 346) If we assume that a, b, and k are independent of each other [that is, $p(a \cap b \cap k) = p(a)p(b)p(k)$], and that $p(a \cap b)$ and $p(k)$ are mutually exclusive [that is, $p[(a \cap b) \cap k] = 0$], Eq. 5.7 simplifies to Eq. 5.8: [28]

$$p(c) = p(a)p(b) + p(k). \tag{5.8}$$

McGuire (1960a, b) constructed 16 syllogisms, one of which read as follows:

Major premise: Any form of recreation that constitutes a serious health menace will be outlawed by the City Health Authority.

Minor premise: The increasing water pollution in this area will make swimming at the local beaches a serious health menace.

Conclusion: Swimming at the local beaches will be outlawed by the City Health Authority.

The 48 propositions constituting the syllogisms were presented in random order, and subjects were asked to "indicate in probabilistic terms the extent of their adherence to the stand taken in each proposition" (McGuire, 1960b, p. 346). These estimates were obtained on graphic scales ranging from *very improbable* (0) to *very probable* (100).

By assuming that $p(k)$ is equal across syllogisms or at least uncorrelated with the conclusion, one may disregard $p(k)$ in a correlational analysis.[29] McGuire's model then implies that the degree of a person's logical consistency is indexed by the correlation between $p(c)$ and $p(a)p(b)$.[30] The results in two of McGuire's (1960a, c) experiments indicated a considerable degree of logical consistency.

---

28. McGuire's basic model states that

$$p(c) = p[(a \cap b) \cup k].$$

Expanding the right-hand side, we get

$$p[(a \cap b) \cup k] = p(a \cap b) + p(k) - p(a \cap b \cap k).$$

Assuming that a, b, and k are mutually exclusive reduces this expression to

$$p(a \cap b) + p(k)$$

and after the assumption is made that a and b are independent,

$$p(c) = p(a)p(b) + p(k).$$

29. Adding a constant to a variable will not alter that variable's correlation with some other variable.

30. Because of the indeterminacy of $p(k)$, exact point predictions of $p(c)$ from $p(a)p(b)$ cannot be made; it can only be stated that $p(c) \geqslant p(a)p(b)$.

McGuire obtained group means for $p(a)$, $p(b)$, and $p(c)$ within each syllogism and computed the correlation between $p(a)p(b)$ and $p(c)$ across syllogisms. In one study (1960a), in which 16 syllogisms were used, the obtained correlation was .48, and in another study (1960c), with eight syllogisms, the correlation was .74.

Similar results were reported by Holt and Watts (1969) and Watts and Holt (1970), who used eight of McGuire's syllogisms in further tests of the model. Correlations were reported by Holt and Watts; the average correlation between predicted and obtained probabilities of the conclusion was .53. Similarly, Dillehay, Insko, and Smith (1966) used all 16 of McGuire's syllogisms and obtained an average correlation of .74 in one study and .54 in a second study.

McGuire (1960a, c) also attempted to investigate the notion that a person's attitudes toward the propositions (i.e., his ratings of their desirability) may influence the degree to which he is logically consistent. His subjects indicated their feelings regarding the desirability of each of the propositions on a five-place scale ranging from *very desirable* to *very undesirable*. A correlation of .40 was found between the mean desirability and probability ratings across 48 propositions (McGuire, 1960a). Other investigators (e.g., Holt and Watts, 1969; Watts and Holt, 1970; Dillehay, Insko, and Smith, 1966) have also found a tendency for subjects to agree more with desirable than with undesirable belief statements. However, as McGuire (1960a) has pointed out, this correlation between perceived likelihood and desires does not necessarily represent cognitive distortion or "wishful thinking." It may simply reflect the (plausible) view that desirable events are indeed more likely to occur than undesirable events.

Therefore, in order to examine the extent to which the desirability of propositions affects logical consistency, McGuire recomputed the correlation between $p(c)$ and $p(a)p(b)$ while statistically controlling for desirability. Specifically, he computed the partial correlation between mean $p(c)$ and mean $p(a)p(b)$ scores across all syllogisms, with the rated desirability of the conclusion and the mean of the desirability ratings of the two premises partialed out. These analyses increased the correlation from .48 to .85 in the study with 16 syllogisms (McGuire, 1960a) and from .74 to .96 in the eight-syllogism study (McGuire, 1960c). Unfortunately, this type of analysis has not been reported in other studies.[31]

McGuire's findings suggest that although there is considerable evidence for probabilistic consistency, inferences may be partially determined by evaluative

31. It has often been argued that wishful thinking is indicated when the discrepancy between predicted and obtained probabilities of the conclusion $[p(c) - p(a)p(b)]$ is larger for syllogisms with desirable conclusions than for those with undesirable conclusions. This argument assumes that $p(k)$ is constant across syllogisms (see note 29). However, the work of Abelson and his associates discussed above (Abelson and Kanouse, 1966; Kanouse, 1971) suggests that this assumption may be misleading. On the basis of their findings it could be argued that different amounts of evidence are required for acceptance of desirable and undesirable conclusions, and thus $p(k)$ would vary across syllogisms.

consistency. As in syllogistic reasoning, however, it has been found (Watts and Holt, 1970) that the relation between probability and desirability is stronger for low-intelligence subjects than for those of high intelligence. A similar result was found by Dillehay, Insko, and Smith (1966) in that the correlation was higher for freshmen and sophomores than for juniors and seniors. Thus it appears that intelligent and more educated individuals are likely to be logically consistent and are less likely to exhibit "wishful thinking."

A Conditional Consistency Model

In addition to the *categorical* statements of logical syllogisms (that is, O is X) it is possible to consider *conditional* statements of the form: If A then B. When such a conditional statement is taken as a major premise, a minor premise can be added and a conclusion can be deduced (cf. D. M. Johnson, 1972, p. 258). Consider, for example, the following propositions.

Major premise: If A is a fish, then A can swim.

Minor premise: A is a fish.

Conclusion: A can swim.

This formulation can be seen as representative of the situations employed by Wyer (1970a) and Wyer and Goldberg (1970). The situation involving reelection of Governor Smith can again serve as an illustration.

Major premise: If Governor Smith is reelected, then state aid to education will be increased—$p(B|A)$.

Minor premise: Governor Smith will be reelected—$p(A)$.

Conclusion: State aid to education will be increased—$p(B)$.

Wyer and Goldberg used the following probability model to describe inferential belief formation in such situations.

$$p(B) = p(B|A)\ p(A) + p(B|\overline{A})p(\overline{A}). \tag{5.9}$$

The terms $p(B)$, $p(A)$, $p(B|A)$, and $p(B|\overline{A})$ were previously defined; $p(\overline{A})$ refers to the prior probability that A is not true and $p(\overline{A}) = 1 - p(A)$.[32] Wyer (1970a) and Wyer and Goldberg (1970) found considerable support for this model. Note that in contrast to McGuire's model of logical consistency, all terms in Eq. 5.9 can be directly estimated. Indeed, the reader will recall that subjects in these studies provided probability estimates for each term in the equation.

32. This model is a special case of the more general expression

$$p(B) = \sum_{i=1}^{n} p(B|X_i)p(X_i).$$

With two mutually exclusive and exhaustive events in X (that is, A and $\overline{A}$) this expression reduces to Eq. 5.9.

Predicted and obtained values of $p(B)$ should directly correspond if subjects make inferences as prescribed by the normative model. In support of this hypothesis, Wyer (1970a) and Wyer and Goldberg (1970) found that predicted and obtained probabilities corresponded very closely (see Fig. 5.5).

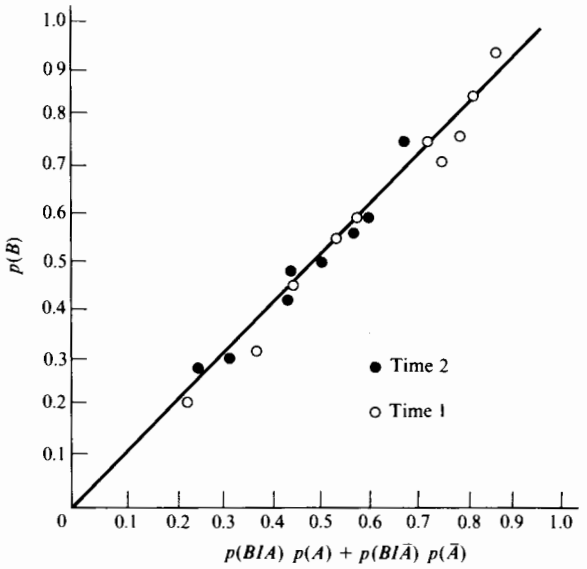

Fig. 5.5 Relation between predicted and obtained probabilities. (Reprinted from Wyer, 1970.)

Wyer and Goldberg further tried to show the correspondence between their model and McGuire's model of logical consistency (see Eq. 5.8). Basically, the $p(a)$, $p(b)$, and $p(c)$ terms in McGuire's model were considered equivalent to $p(A)$, $p(B|A)$, and $p(B)$, respectively, by Wyer and Goldberg. If McGuire's terms are substituted in Eq. 5.9, the following expression results.

$$p(c) = p(a)p(b) + p(c|\bar{a})p(\bar{a}). \tag{5.10}$$

Comparing this expression with McGuire's model in Eq. 5.8, we can see that the two models are identical if $p(k)$ is set equal to $p(c|\bar{a})p(\bar{a})$. Although recognizing that $p(k)$ is not equivalent to $p(c|\bar{a})p(\bar{a})$, Wyer and Goldberg (1970) argued that the latter term can provide an estimate of $p(k)$, thus permitting a more precise test of McGuire's model.

In conclusion, it is apparent that subjective probabilities or beliefs are related to one another in systematic ways and that various normative probability models can provide fairly adequate descriptions of actual inference processes. McGuire (1960a) and Wyer and Goldberg (1970) also attempted to use their models to predict change in beliefs produced by persuasive communications. Clearly, according to these models, a change in one probability or belief is likely to

produce changes in related beliefs, and these changes can be predicted from the model. However, the amount of change will depend on the particular interrelations existing among the various beliefs. These issues will be discussed at length in Chapter 9, which deals with principles of change.

It is important to note that, depending on the assumptions made, different models can be developed to deal with the same issues. We have already seen how the introduction of conditional probabilities by Wyer and Goldberg produced a model that, although similar to McGuire's, did not require an unspecified $p(k)$ term. Similarly, Warr and Smith (1970) compared the predictive power of six different probability models dealing with trait inferences. These models were designed to predict the conditional probability that a person has Trait A, given that he has Traits B and C. Although these models involved somewhat different assumptions, they were all found to predict inferences with considerable accuracy —another indication that trait inferences are based on probabilistic consistency. Note that any of the models used by Warr and Smith (1970) could be applied to an analysis of syllogistic reasoning since they deal with the effect of two pieces of information (e.g., two premises) on a third (e.g., a conclusion).

Bayes's Theorem

The probability models discussed thus far deal primarily with the interrelations among beliefs; when applied to belief change, they deal with the effect of change in one belief on changes in related beliefs. These models, however, do not address themselves to the question as to how the initial change is brought about. Bayes's theorem deals specifically with this problem; that is, it is concerned with the formation or revision of beliefs in light of new information. Again note that Bayes's theorem is a normative model in that it describes optimal revisions in probabilities; that is, it describes how probabilities should change if the available information is properly utilized. The theorem deals with revision of beliefs or hypotheses (H) on the basis of new information or data (D). In its simplest form, Bayes's theorem is expressed in Eq. 5.11.[33]

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}. \tag{5.11}$$

Consider a person who sees a stranger at a cocktail party. Without any further information, he might believe that there is a .30 probability that the stranger is a liberal. That is, on the basis of his past experience he might believe that approximately 30 out of 100 people are liberals. He now has a short conversation with the stranger, in the course of which he learns that the stranger voted for Goldwater in 1964. Bayes's theorem specifies how this new information should influence the person's belief that the stranger is a liberal—that is, what the posterior

33. The reader may note that this formulation can be derived from Eq. 5.3 by dividing both sides of the equation by $p(B)$. A discussion of Bayes's theorem and its implications can be found in W. Edwards, Lindman, and Savage (1963).

probability concerning the stranger's liberalness should be, given the new information that he voted for Goldwater. According to Eq. 5.11, three probabilities must be known to answer this question.

1. $p(D|H)$ The probability that a person voted for Goldwater, given that he was a liberal
2. $p(H)$ The prior probability that a person is a liberal
3. $p(D)$ The prior probability that a person voted for Goldwater

Let us assume that $p(D|H) = .10$ and that $p(D) = .40$; since we already know that $p(H) = .30$, we can compute $p(H|D)$, as follows:

$$p(H|D) = \frac{.10\ (.40)}{.30} = .13.$$

Thus, according to Bayes's theorem, the new information should reduce the person's subjective probability that the stranger is a liberal from .30 to .13. This predicted revision in subjective probability could now be compared with the person's actual revision to assess the degree to which the normative model is descriptive of actual inferences.

Bayes's theorem can also be written

$$p(H|D) = \frac{p(D|H)p(H)}{p(D|H)p(H) + p(D|\bar{H})p(\bar{H})}, \tag{5.12}$$

since $p(D) = p(D|H)p(H) + p(D|\bar{H})p(\bar{H})$. [This equation was adopted by Wyer and Goldberg (1970) as their model of conditional consistency; see Eq. 5.9.] One interesting implication of Bayes's theorem as presented in Eq. 5.12 is that no revision in belief should occur when $p(D|H) = p(D|\bar{H})$. That is, when the datum is equally likely whether or not the hypothesis is true, it should have no effect on the hypothesis. To return to our earlier example, if liberals and nonliberals were equally likely to have voted for Goldwater, the datum that the stranger voted for Goldwater would provide no information as to how liberal he is. In fact, when $p(D|H) = p(D|\bar{H})$, these terms drop out of Eq. 5.12, which then becomes

$$p(H|D) = \frac{p(H)}{p(H) + p(\bar{H})}.$$

However, $p(H) + p(\bar{H}) = 1$, so that $p(H|D) = p(H)$, and H and D are thus by definition independent (see Eq. 5.5). It follows that the datum D should have no effect on H when $p(D|H) = p(D|\bar{H})$.

When the difference between these two conditional probabilities is positive, i.e., when $p(D|H) - p(D|\bar{H}) > 0$, an upward revision in probability is expected. Conversely, when $p(D|H) - p(D|\bar{H}) < 0$, a downward revision should result. The greater the difference between $p(D|H)$ and $p(D|\bar{H})$, the more impact the datum in question should have on the posterior probability.

Clearly, a person may have several alternative hypotheses concerning a given situation. To return to our previous example, the stranger might be a liberal, moderate, or conservative, and the observer might believe that the corresponding probabilities of these hypotheses are .30, .50, and .20, respectively. The information that the stranger voted for Goldwater might lead to revisions in all these probabilities. Eq. 5.12 assumes two alternative hypotheses, $p(H)$ and $p(\overline{H})$. With reference to our example, $p(H)$ is the probability that the stranger is a liberal and $p(\overline{H})$ that he is not a liberal (i.e., a moderate or a conservative).

In many situations it is useful to consider two alternative hypotheses, $p(H)$ and $p(\overline{H})$, and Bayes's theorem can be stated with respect to each, as follows:

1. $p(H|D) = \dfrac{p(D|H)p(H)}{p(D)}$.

2. $p(\overline{H}|D) = \dfrac{p(D|\overline{H})p(\overline{H})}{p(D)}$.

A computationally convenient form of Bayes's theorem for two hypotheses is provided when the first of these equations is divided by the second:

$$\frac{p(H|D)}{p(\overline{H}|D)} = \frac{p(D|H)}{p(D|\overline{H})} \cdot \frac{p(H)}{p(\overline{H})} \tag{5.13}$$

In Eq. 5.13

$$\frac{p(H|D)}{p(\overline{H}|D)}$$

is the posterior odds with respect to Hypothesis H, that is, the extent to which Hypothesis H is more or less likely than Hypothesis $\overline{H}$. (Odds are interpreted as in betting situations.) Similarly,

$$\frac{p(H)}{p(\overline{H})}$$

is the prior odds with respect to Hypothesis H. Finally,

$$\frac{p(D|H)}{p(D|\overline{H})}$$

is known as the *likelihood ratio,* which is an index of the *diagnosticity* of the datum in question; that is, it indicates the degree to which the datum favors one hypothesis over the other. Equation 5.13 is often written more simply as

$$\Omega_1 = \text{LR}\ \Omega_0 \tag{5.14}$$

where Ω_1 is the posterior odds, Ω_0 is the prior odds, and LR is the likelihood ratio.

As Eq. 5.13 shows, the likelihood ratio comprises the two conditional probabilities considered earlier with respect to Eq. 5.12, that is, $p(D|H)$ and $p(D|\overline{H})$. Again, when these two terms are equal, LR = 1 and $\Omega_1 = \Omega_0$, and thus no revision in probability takes place. When $p(D|H)$ exceeds $p(D|\overline{H})$, the likelihood

ratio becomes greater than 1, and the odds are expected to be revised in favor of Hypothesis H. Conversely, when $p(D|H) < p(D|\overline{H})$, LR < 1, and the odds are expected to be revised in favor of Hypothesis $\overline{H}$.

Suppose that prior to receipt of any information about the stranger in our previous example, the odds were 4:6 that he was a Republican rather than a Democrat. Further, assume that LR = 20:1; that is, the chances are 20 to 1 that a person who is a Republican rather than a Democrat would have voted for Goldwater. According to Bayes's theorem, the posterior odds should be 40 to 1 that the stranger is a Republican rather than a Democrat. Since the likelihood ratio greatly exceeded unity, the revision from Ω_0 to Ω_1 was considerable and in the positive direction.

Bayes's theorem can be applied sequentially for understanding the impact of several data. The posterior probability resulting from the first item of information is used as the prior probability for the second item of information, etc. Depending on the particular assumptions made, Bayes's theorem specifies whether the final posterior belief is influenced by the order in which these new items of information are received. When conditional independence between items of information is assumed, Bayes's theorem predicts that order does not affect the final posterior belief. When conditional independence is not assumed, Bayes's theorem suggests that order effects can strongly influence the final posterior odds. Thus, Bayes's theorem prescribes optimal aggregation of several items of information.[34]

A large number of studies have shown that Bayes's theorem is a reasonably good descriptive model of human information processing (see Peterson and Beach, 1967, for a good review of this literature). Revisions in beliefs tend to be probabilistically consistent and to correspond quite well to the prescriptions of Bayes's theorem.[35] In other words, there is a strong tendency not only for beliefs to be internally consistent but also for people to revise their beliefs in an orderly fashion as a result of new information. Despite this internal consistency of belief structures, however, many studies have found a tendency for subjects to be "conservative" information processors. That is to say, although "opinion change is very orderly, and usually proportional to numbers calculated from Bayes' theorem, . . . it is insufficient in amount" (Edwards, 1968, p. 18). Indeed, this tendency toward conservatism has been the focus of much of the research related to Bayes's theorem (see Edwards, 1968, and Slovic and Lichtenstein, 1971, for reviews of this literature).

The most common Bayesian experiment examines revision of beliefs about the characteristics of a population in light of a sample taken from that population.

---

34. The final posterior odds, given n items of conditionally independent information, are

$$\Omega_n = \prod_{i=1}^{n} \mathrm{LR}_i \, \Omega_0.$$

35. However, see Kahneman and Tversky (1972) for a review of some evidence indicating that individuals may not be Bayesian information processors.

Thus a subject may be told that a bag contains either 70 percent blue and 30 percent red poker chips, or 30 percent blue and 70 percent red chips. He may then be shown a sample of, say, six blue and four red poker chips, taken at random (with replacement) from the bag. The subject estimates (either after each poker chip is drawn from the bag or at the end) the likelihood that the sample was taken from the predominantly blue bag. Since the prior probability of this event is .50, and since the likelihood ratio can be objectively determined in this situation, it is possible to compute the posterior probability by using Bayes's theorem. In this example, the posterior probability is approximately .845.[36] In comparison with this prediction, subjects tend to exhibit conservatism by estimating a somewhat lower probability.

It has been argued that this effect may be due either to the subject's *misperception* of the diagnostic value of a given datum [i.e., of the likelihood ratio $p(D|H)/p(D|\overline{H})$] or to his *misaggregation* of the data. That is, a person's subjective probabilities prior to receiving a new item of information may not be accurate, and hence his inference will not correspond to that prescribed by Bayes's theorem. Alternatively, even though the person's subjective probabilities concerning the impact of any given item of information may be accurate, he may fail to aggregate several such items in accordance with the normative model. We can see that this distinction is essentially the same as our earlier distinction between *obtaining* knowledge or information and *using* that knowledge appropriately.

The question of conservatism is concerned with the degree to which a person's inferences are *accurate,* rather than the degree to which they follow consistently from his perception of prior and conditional probabilities. As we have noted earlier, there is abundant evidence to indicate that an individual's belief structure is internally consistent and that it tends to be compatible with Bayes's theorem. Thus a person may misperceive the objective probabilities, but when his subjective probabilities are considered, Bayes's theorem allows quite accurate predictions. However, some evidence appears to indicate that subjects tend to misaggregate information; that is, their estimates of posterior probabilities tend to be conservative even in comparison with Bayesian predictions based on subjective probabilities provided by the subjects themselves (e.g., Peterson and Miller, 1965; Phillips, 1966). This apparent tendency for subjects to misaggregate data seems to be attributable to a subject's reluctance to give an extreme response early in the stimulus sequence for fear of arriving at the end of the probability scale before the last datum is presented. Indeed, investigators have found that the "misaggregation" effect disappears and conservatism is reduced when subjects are asked to

---

36. The equations are

$$\Omega_{10} = \Pi \, \mathrm{LR} \, \Omega_0 = (7/3)^6 \cdot (3/7)^4 \cdot 1 = 5.44,$$

and

$$p(H|D) = \frac{5.44}{6.44} = .845.$$

(See note 33.)

provide estimates of odds rather than probabilities (DuCharme, 1970; Phillips and Edwards, 1966).

W. Edwards (1962) attempted to develop a strategy for aiding the decision maker such that conservatism would be eliminated. Since he assumed that conservatism was largely due to misaggregation, he proposed that experts provide estimates of likelihood ratios (and prior probabilities), and that Bayes's theorem be used to generate predictions of posterior probabilities. This procedure is quite similar to Goldberg's (1970) "bootstrapping" strategy based on research in cue utilization, which we mentioned previously. Although our discussion above suggests that misaggregation may not be a major contributor to conservatism, Edwards's strategy would nevertheless outpredict man. That is, as Goldberg suggested, the normative model operates more consistently than man and thus eliminates various sources of human error or response bias.

Very few applications of Bayes's theorem can be found in attitude research. Nevertheless, the few studies reported have dealt with such diverse areas as trait inferences (C. R. Peterson *et al.*, 1965), impression formation (Cohen, 1973), perception of interpersonal relations (McNeel and Messick, 1970), attribution processes (Ajzen, 1971a), attitude-behavior relations (Fishbein and Ajzen, 1974), and reactions to disconfirmed expectancies (Brickman, 1972). One study of relevance to our discussion of inferential belief formation is Brickman's (1972) attempt to show that apparently irrational reactions to disconfirmed expectations may actually be quite rational revisions in beliefs; i.e., the reactions may be consistent with prescriptions of Bayes's theorem. Suppose that a person expected to get an A on an exam and received a hint that he had actually obtained a C. In comparison with a person who expected to get a C, the person who expected an A would be reluctant to accept this information; that is, his posterior probability of having a C would be relatively low. Although the apparent reluctance to accept disconfirming or unexpected information may appear to be an irrational attempt to avoid inconsistency or tension, Bayes's theorem suggests that the observed differences in posterior probabilities may merely reflect different prior probabilities. That is, a person will require more evidence to conclude with confidence that his grade is a C when he expected an A than when he expected a C.

Although Brickman's data analysis is open to criticism, his results appear to provide tentative support for these notions. Following an examination, students volunteered to learn their grades in the context of an experiment. Each student received a set of cards. All students were told that each card contained the letter A, B, C, or D and that the letter appearing most frequently in the set represented the grade for that student. They provided estimates of prior probabilities by distributing 100 points among the four letter grades. After viewing each card, each student revised his estimate by redistributing the 100 points. Each of the letters A, B, and C appeared most frequently for one third of the subjects selected at random. Thus the information was consistent with some subjects' prior probabilities and inconsistent with others'. Subjects could view as many cards as they

desired. At the end of the experiment they indicated their certainty about their grades on an 11-point scale.

Consistent with expectations, subjects who received discrepant information were less certain about their grades than were subjects who received information consistent with their expectations. Since certainty can be viewed as a measure of subjective probability (see Chapter 3), this finding supports the notion that unexpected information leads to lower posterior probabilities than does expected information. However, comparing trial-by-trial *revisions* in subjective probabilities with optimal revisions prescribed by Bayes's theorem, Brickman found that unexpected information produced, on the average, smaller deviations than did expected information (but always in the direction of conservatism). Thus there was no evidence that disconfirmation of expectancies resulted in reluctance to accept the information; on the contrary, it appeared that unexpected information was used more rationally than was expected information. Unfortunately, the latter conclusion is open to question since alternative interpretations cannot be ruled out.[37] This study nevertheless provides an interesting application of Bayes's theorem, and the results suggest again that inferential belief formation tends to follow laws of probabilistic rather than evaluative consistency or tendencies toward wishful thinking.

In conclusion, Bayes's theorem specifies optimal revision of beliefs in light of new information, and it thus constitutes a model of a "rational" person whose inferences are unaffected by his desires or by other extraneous considerations. Available evidence indicates that the model provides a fairly accurate description of actual inference processes. Bear in mind, however, that at the present time Bayes's theorem has been applied only in a limited range of situations, and it is possible that greater deviations from the model will be observed in other situations.

Bayes's theorem deals with the impact of a new item of information on one or more beliefs or hypotheses. With two hypotheses (e.g., O is X versus O is not X) the likelihood ratio described earlier represents the information's *diagnosticity* with respect to the two alternative hypotheses. That is, the likelihood ratio indexes the datum's relative probability of occurrence under the two alternative hypotheses; the more the likelihood ratio deviates from unity (in either direction), the greater the information's diagnosticity. According to the model, the amount of change in a given belief is a function of the diagnostic value of the new item of information.[38]

---

37. Despite the author's claim to the contrary, the observed discrepancies between predicted and obtained revisions may be influenced by prior probabilities and may correlate with number of trials. Either of these factors could explain the greater conservatism obtained with consistent than with inconsistent information.

38. This can be seen most clearly when Bayes's theorem is stated in logarithmic form: $\log \Omega_1 - \log \Omega_0 = \log \text{LR}$ (see Eq. 5.14). In this form, the amount of revision is equivalent to log LR, and it is independent of the prior odds.

The theorem itself, however, has little to say about the degree to which some new item of information has diagnostic value for a given inference. This value has to be determined on the basis of other considerations. In some situations, such as the poker-chip guessing task or in Brickman's (1972) study described above, it is possible to compute objective likelihood ratios; but in most situations of interest to the social psychologist, this cannot be done. When making an inference about a person's religiosity, for example, one would find it very difficult to determine the exact (objective) diagnosticity of the fact that the person in question was married in church. At best, we could ask our respondent to make a *subjective* estimate of the likelihood ratio under consideration and use it to predict his inference.

This is not to say, however, that it is impossible to develop some general ideas or theories about the factors that influence diagnosticity of certain kinds of information. Indeed, the different theories of attribution discussed in Chapter 2 can be viewed as attempts to specify some of those factors with respect to one kind of information, namely, information about a person's behavior. In the following section we will try to show that most of the factors which have been stipulated to influence causal attributions would also be expected to affect the diagnosticity of information on which attributions are based.

ATTRIBUTION PROCESSES

A person who observes the behavior of an actor can usually find many alternative explanations to account for the behavior; that is, he can attribute the observed behavior to different causal factors. Consider, for example, an actress who laughs at her husband's joke. In an attempt to explain this behavior, an observer might consider several hypotheses, such as: the joke is funny; the actress wants to please her husband; the husband knows how to tell jokes, etc. Unlike the poker-chip guessing task in a typical Bayesian experiment, the attribution situation involves a variety of possible hypotheses that need not be mutually exclusive. For each hypothesis the observer must determine how likely it is that the behavior in question would have been produced by, or could have been attributed to, the hypothesized causal factor.

Let B stand for the observed behavior (the datum) and H for the belief or hypothesis invoked to explain the behavior. The extent to which the behavior is viewed as produced by a given factor is represented by the likelihood ratio: $p(B|H)/p(B|\overline{H})$. The greater this ratio, the more likely it is that the behavior would have been performed if the hypothesized causal factor was present, i.e., if the hypothesis was correct. To return to our example, when the actress is viewed as likely to have laughed at her husband's joke if she wanted to please him but not otherwise, her desire to please her husband constitutes a likely explanation for her behavior.

We can see that causal attribution is here interpreted as corresponding to the likelihood ratio in Bayes's theorem. This interpretation has important implications for the effects of causal attributions on the observer's beliefs. Each alternative hypothesis constitutes at once a possible causal explanation for the behavioral

event and a belief about one aspect of the total situation. For instance, the hypothesis that the joke is funny can be invoked to explain why the actress laughed at her husband's joke; at the same time it constitutes a belief about a property of the joke. The observer's posterior probability that the joke is funny has frequently been used as a measure of attribution. However, Bayes's theorem suggests that causal attribution (i.e., the likelihood ratio) corresponds to a *revision* in the belief that a given factor is present, rather than to the posterior odds in favor of that belief.

That causal attribution involves revisions from prior to posterior odds can be seen most clearly in the case of an irrelevant hypothesis. Consider the rather unlikely hypothesis that the actress laughed at her husband's joke *because* she was a brunette. Clearly, this search for a causal explanation cannot be equated with the posterior probability that the actress is a brunette. Let us assume that she indeed has brown hair. Following observation of her behavior, the observer will have a high subjective probability that the actress is a brunette (and a low subjective probability that she is not). This does not mean, however, that he will view her hair color as a satisfactory explanation for her behavior. According to Bayes's theorem, this explanation will be invoked only to the extent that the observer has *raised* his odds that the actress is a brunette *as a result* of observing her behavior. In our example, the observer's belief about the color of the actress's hair is not likely to have been influenced by the fact that she laughed at her husband's joke. It follows that this belief will not be viewed as a likely explanation for the behavior.

Most research on attribution processes has examined the effects of variations in experimental conditions on causal attributions. In terms of Bayes's theorem, one can argue that these experimental variations influence the diagnostic value of the observed behavior. Many studies have obtained direct estimates of the extent to which certain hypotheses were invoked as causal explanations for an observed behavioral event (e.g., Feather, 1969; Feather and Simon, 1971a,b; Weiner and Kukla, 1970; Frieze and Weiner, 1971; McArthur, 1972). Other studies have measured one or more of the observer's beliefs—usually about the actor's dispositions (e.g., Steiner and Field, 1960; Jones, Davis, and Gergen, 1961; Jones and Harris, 1967; Jones, Worchel, Goethals, and Grumet, 1971; Ajzen, 1971a). Most of the latter studies, however, have assessed only posterior probabilities, rather than revisions from prior to posterior odds. On the assumption that the prior odds (or probabilities) are the same in all experimental conditions, obtained differences in posterior odds should also be a direct function of the likelihood ratio, and the posterior odds or probabilities can be used as a measure of attribution. Unfortunately, posterior probabilities have sometimes been used as a measure of attribution even when prior probabilities could not be expected to be the same across experimental conditions, and from our point of view, such posterior-probability measures provide little information about attribution processes.

In conclusion, attribution theory deals with the perceived likelihood of alter-

native causal factors as *explanations* of observed behavior. The extent to which a potential causal factor is viewed as responsible for the behavior corresponds to the likelihood ratio in Bayes's theorem, where the behavior serves as the datum and the proposed explanation as the hypothesis. According to the Bayesian model, such causal attribution is equivalent to revisions in the hypotheses or beliefs under consideration.

Causal Attribution

An actor's behavior can usually be attributed to a large number of causal factors. If the only information available to the observer is the fact that the actor has performed the behavior in question, causal attribution may prove quite difficult. Frequently, however, the observer will have some additional information about the actor, the behavior, the conditions under which the behavior was performed, and the effects it produced. We shall now examine the effects of such additional information on the perceived likelihood of alternative causal explanations.

Effects of Consistency

In Chapter 2 we discussed several factors that are assumed to influence causal attribution. On the basis of Heider's (1958) theorizing, Kelley (1967, 1973) proposed two principles to account for causal attribution of an observed effect (or behavior). According to the *covariation principle,* "An effect is attributed to the one of its possible causes with which, over time, it covaries" (Kelley, 1973, p. 108). The *discounting principle* states that "the role of a given cause in producing a given effect is discounted if other plausible causes are also present" (Kelley, 1973, p. 113).

Consistency across occasions, objects, and actors. The covariation principle is closely tied to the notion of consistency. Kelley's analysis deals with the degree to which the behavior in question is displayed consistently on *different occasions* ("consistency"), toward *different objects* ("distinctiveness"), and by *different actors* ("consensus").

McArthur's (1972) study illustrates these three aspects of consistency. In a direct test of Kelley's theory, McArthur described 16 hypothetical responses of different actors, such as "Ralph trips over Joan's feet while dancing," "Sue is afraid of the dog," and "Bill thinks his teacher is unfair." Consistency was manipulated by providing additional items of information. For example, to create high (or low) consistency across occasions, subjects were told that "In the past, Ralph has almost always (never) tripped over Joan's feet." High and low consistency across objects are illustrated, respectively, by "Ralph also trips over almost every other partner's feet" and "Ralph does not trip over almost any other partner's feet." Finally, high and low consistency across actors took the form of "Almost everyone (hardly anyone) else who dances with Joan trips over her feet." For each of the 16 hypothetical situations, the three kinds of consistency were varied in a $2 \times 2 \times 2$ analysis of variance design.

Subjects indicated which of four alternatives had caused the actor's behavior: (1) something about the actor; (2) something about the object; (3) something about the particular circumstances; or (4) some combination of the first three alternatives. Note that the dependent measure of attribution in this study was a direct indication of the causal factor held responsible for the actor's behavior. A subject's response can be viewed essentially as an indication of the behavior's likelihood in the presence of a given causal factor (e.g., something about the object) relative to its likelihood under the alternative hypotheses. The response can therefore be interpreted as roughly corresponding to the likelihood ratio, $p(B|H)/p(B|\overline{H})$.

A Bayesian analysis of the expected effects is consistent with Kelley's (1973) theorizing. Consider the hypothesis that something about the actor is responsible for his behavior. The likelihood ratio for this hypothesis should increase with consistency across occasions and objects, and it should decrease with consistency across actors. To illustrate, Ralph is more likely to trip over Joan's feet when he possesses an appropriate disposition (such as clumsiness) than when he does not; that is, $p(B|H) > p(B|\overline{H})$. Each additional occasion on which Ralph trips over Joan's feet serves to increase the likelihood ratio. A similar argument can be made for repeated performance of the behavior with different partners. In contrast, information that other actors also trip over Joan's feet should increase $p(B|\overline{H})$, the probability that Ralph will trip over her feet even if he possesses no attribute that would predispose him to do so; the likelihood ratio should therefore decrease with consistency across actors.

Similar analyses with respect to the hypotheses that Ralph's behavior is due to something about Joan (the object) or about the particular circumstances lead to the following predictions: The likelihood ratio and hence causal attribution to the object should increase with consistency across occasions and actors but decrease with consistency across objects. For attribution to circumstances, the likelihood ratio should decrease as a function of consistency across actors, objects, and occasions.

McArthur (1972) tested these expectations by analyzing the frequencies with which the behavior was attributed to actor, object, and circumstances. Generally speaking, her results lent support to the predictions above. Additional empirical evidence for the predicted effects of consistency has been reported by Kelley and Stahelski (1970) and by Frieze and Weiner (1971).

Attributions to actor, object, or circumstances, however, do not provide the only possible explanations for the actor's behavior. As noted, McArthur's subjects could also attribute the behavior to some combination of those factors. By far the most frequent combinational hypothesis concerned the interaction between actor and object. Certain types of information led subjects to hypothesize that the behavior must have been caused by something peculiar to the particular combination of actor and object. This hypothesis was invoked, for instance, when there was high consistency across occasions but low consistency across objects and actors.

To see how such information would influence the likelihood ratio for the actor-object interaction, let us again consider Ralph, who trips over Joan's feet. When few other people trip over Joan's feet, when Ralph trips over few other partners' feet, but when he trips over Joan's feet time and time again, no single factor can account for the behavior. However, the probability of the behavior should be high if one assumes that the particular actor-object combination is at fault, leading to a high likelihood ratio for the combinational hypothesis.

Consistency across behaviors. We have seen that variations in consistency across occasions, objects, and actors are expected to influence the likelihood ratio in a manner compatible with Kelley's (1967, 1973) analysis of causal attribution. However, a Bayesian analysis suggests that the likelihood ratio may also be affected by the degree of consistency along other dimensions. Of particular importance is the consistency among *different behaviors* performed by the actor. It has often been argued that a person's attitudes (and other dispositions) are evidenced by consistency in his responses toward the attitude object (response-response consistency). In a similar manner it can be proposed that an observer attributes an actor's behavior to a given causal factor to the extent that other behaviors performed by the same actor are consistent with the proposed explanation.

Consider, for example, the hypothesis that Ralph tripped over Joan's feet because he is clumsy. The perceived likelihood of this explanation should increase if Ralph was also observed to spill his coffee, to drop his books, to burn his fingers while lighting a cigarette, etc. In terms of the likelihood ratio, the subjective probability of the totality of these behavioral events will be higher under the hypothesis that Ralph is clumsy than under the hypothesis that he is not. The total set of behaviors, including the tripping over a dance partner's feet, is therefore likely to be attributed to the actor's clumsiness.

Effects of sample size. The notion of consistency discussed so far refers to the *proportion* of consistent events. Specifically, we have dealt with the proportion of occasions on which a given behavior is performed, the proportion of objects toward which it is displayed, the proportion of actors who perform the same behavior, and the proportion of behaviors that are consistent with the hypothesis. The notion of consistency corresponds closely to the proportion of red and blue poker chips in a typical Bayesian experiment. The greater the proportion of red chips, the more likely it is that the sample was taken from the predominantly red bag and the less likely it is that it was taken from the predominantly blue bag. That is, the greater the proportion of red chips, the more the likelihood ratio will differ from unity. However, when the sample of chips is inconsistent (i.e., when an equal number of red and blue chips are sampled), it is about as likely that the sample was drawn from the red bag as from the blue bag and the likelihood ratio will be close to unity.

Work on Bayes's theorem suggests that when proportion is held constant, diagnosticity and hence causal attribution should increase with sample size, i.e., with the *number* of events that have been observed. Suppose that a student of

average ability would pass about 50 percent of a series of tests. Information that a certain student has passed three out of four tests has considerably lower diagnostic value for inferring his ability than information that he has passed 30 out of 40 tests, although his performance is 75 percent consistent in both instances. Stated differently, it is not unlikely that a student of average ability would pass one more test than would be expected by chance (i.e., 3 out of 4) but a student who passes 10 more tests than expected (i.e., 30 out of 40) is likely to be perceived as having above average ability. Research in the poker-chip guessing situation has shown that the perceived likelihood ratio (and revision in subjective probabilities) increases with sample size (e.g., DuCharme and Peterson, 1969). Kahneman and Tversky (1972), however, have reported data suggesting that subjects may disregard sample size and base their posterior-probability judgments on the most salient characteristics of the sample, such as its proportion or mean. These conflicting findings have yet to be reconciled.

Attributions based on success and failure. In his treatment of attribution processes, Heider (1958) presented a rather detailed analysis of the factors that may be invoked to explain success or failure on a task. Adopting Heider's notions, Weiner and his associates (Weiner and Kukla, 1970; Weiner *et al.,* 1971; Frieze and Weiner, 1971) proposed four factors that may serve to explain success or failure on a task: ability, effort, task difficulty, and luck. The first two factors are internal, the last two external. Further, ability and task difficulty are relatively stable factors, and effort and luck are relatively unstable factors. This classification can be seen in Table 5.8.

Table 5.8 Perceived Determinants of Success and Failure (from Weiner *et al.,* 1971)

	Internal	External
Stable	Ability	Task difficulty
Unstable	Effort	Luck

A systematic investigation of the effects of consistency and other variables on causal explanation of success and failure has been reported by Frieze and Weiner (1971). They found that in comparison with failure, success on a task tended to be attributed to greater ability, effort, luck, and easiness of the task.[39] Obviously, these findings are quite consistent with a Bayesian analysis since success is more likely when these factors are present than when they are absent.

39. The dependent measure in this study constitutes a relatively direct measure of the likelihood ratio since subjects were asked to rate the likelihood of success or failure, given each of the four alternative hypotheses.

Of greater interest, causal attribution was also affected by consistency of an actor's present success or failure with his past performance on the same task, with his performance on similar tasks, and with the performance of other actors. The results were again in line with a Bayesian analysis. Consistent performance on the same or similar tasks increased attribution to stable factors whereas inconsistent performance increased attribution to unstable factors. Thus the success of an actor who in the past had always performed the same task successfully was attributed mainly to his ability and to the easiness of the task whereas the success of an actor who had always failed on the task was attributed to good luck. In terms of likelihood ratios, an actor's consistent success is more likely, given ability or an easy task than given lack of ability or a difficult task. However, success following failure is more likely, given good than bad luck. A Bayesian analysis also accounts for the finding that consistent performance across *actors* increased attribution to the task and reduced attribution to ability, effort, and luck.

In sum, the effects on causal attribution predicted on the basis of Kelley's (1967, 1973) covariation or consistency principle can also be derived from a consideration of the likelihood ratio in Bayes's theorem. The Bayesian analysis, moreover, suggests a number of related factors of potential importance, such as behavioral consistency and sample size.

Effects of Multiple Plausible Causes

Kelley's *discounting principle* deals with the effects of multiple plausible causes on the likelihood that a given causal factor will be invoked to explain the actor's behavior. A Bayesian analysis suggests that alternative plausible hypotheses will tend to reduce any given explanation's likelihood ratio by raising $p(B|\overline{H})$, the probability of the behavior in the absence of the causal factor under consideration.

A case in point is Bem's (1965, 1967) attributional analysis of dissonance phenomena, particularly with respect to counterattitudinal behavior under insufficient justification. Bem used the "interpersonal replication procedure," in which observers are given descriptions of one or another condition in a dissonance experiment and are asked to infer the actual subject's responses. In one study, for example, this procedure was applied to the Festinger and Carlsmith (1959) experiment on forced compliance. Observers were told that the actor who had just participated in an experiment agreed to tell and actually did tell a waiting subject that the experiment had been interesting. In one condition, the actor was said to have been offered \$20, in another condition \$1, for performance of this behavior. The observer's task was to infer the original subject's attitude toward the experiment. Bem found that his observers attributed a more favorable attitude to the subject in the low (\$1) than in the high (\$20) reward condition.

The dependent variable in this study can be viewed as the observer's posterior probability that the actor had a favorable attitude toward the experiment, given that he said the experiment was interesting. Since it seems reasonable to assume that the actor's perceived attitude prior to his telling the waiting subject that the

experiment was interesting did not differ between conditions, this posterior probability can serve as a measure of attribution.

To account for Bem's findings, consider the likelihood ratio $p(B|A+)/p(B|\overline{A+})$, where B is the behavior of telling a waiting subject that the experiment was interesting, and $A+$ is the hypothesis that the actor had a favorable attitude toward the experiment. With other things equal, if the actor had a favorable attitude, he should be more likely to perform the behavior than if he did not. However, $p(B|\overline{A+})$ should increase with monetary reward offered for the behavior, producing a lower likelihood ratio under high-reward conditions. It follows that the actor's behavior should be attributed to a favorable attitude in the \$1 condition but less so in the \$20 condition. The same argument can be made for any other factor that serves to justify and hence to provide an alternative explanation for the actor's behavior, such as social pressure or threat of punishment for noncompliance.

Attribution of Dispositions

Internal Versus External Attributions

We have discussed variables that influence the extent to which an actor's behavior is attributed to alternative causal factors. It is often convenient to distinguish between factors *internal* to the actor, such as his stable dispositions or temporary moods and desires, and *external* factors residing in the object of his behavior or the environment. Inspired by Jones and Davis's (1965) monograph on attribution processes in person perception, much recent research has dealt with the conditions under which an observed behavior will provide the basis for a dispositional attribution. The question of interest is whether observation of a behavior will lead to the inference that the actor possesses a certain disposition (an internal factor) or to the inference that some external factor must have been present. To simplify exposition, assume that the observed behavior (B) can be attributed either to a dispositional internal factor (D) or to a nondispositional external factor ($\overline{D}$); that is, $p(\overline{D}) = 1 - p(D)$. Bayes's theorem with respect to these two alternative hypotheses can then be written as follows:

$$\frac{p(D|B)}{p(\overline{D}|B)} = \frac{p(B|D)}{p(B|\overline{D})} \cdot \frac{p(D)}{p(\overline{D})}. \tag{5.15}$$

Any factor influencing the likelihood ratio $p(B|D)/p(B|\overline{D})$ should have an effect on dispositional attribution; the greater this likelihood ratio, the more likely it is that a dispositional attribution will be made.

Based on Heider's (1958) analysis of personal causality, Jones and Davis (1965) proposed that a disposition can be attributed to an actor with confidence to the extent that the actor is perceived to have behaved intentionally. Jones and Davis therefore concentrated their analysis on the factors that influence *attribution of intention*. Following Heider, they suggested that the effects or outcomes

produced by the actor's behavior are powerful determinants of dispositional attribution. An observer will attribute intentionality if he believes that the actor was aware that his actions would produce the observed effects and if he believes the effects were due to the actor's ability rather than chance factors.

Two major variables have been investigated in this context: the actor's perceived decision freedom and the perceived desirability or utility of his behavior. Generally speaking, dispositional attribution is expected to increase with perceived freedom of choice and to decrease with behavioral utility.

Perceived decision freedom. Several investigators have proposed that dispositional attribution will increase to the extent that the actor is perceived to have behaved under high freedom of choice. In an early study, Steiner and Field (1960) led subjects to believe that a confederate took a pro-segregation stand in a group discussion either under assignment to that role by the experimenter (low decision freedom) or of his own free choice (high decision freedom). Subjects were more confident in attributing a pro-segregation stand to the confederate under high than under low freedom of choice.[40]

The prediction of stronger attribution under high decision freedom also follows from a Bayesian analysis. It stands to reason that a behavior performed under low freedom of choice has little diagnostic value; that is, the behavior (B) is as likely to be performed with or without the disposition (D) in question. Under high freedom of choice, however, the behavior is more likely to be performed when the actor has the appropriate disposition than when he does not. To put this more formally, the likelihood ratio $p(B|D)/p(B|\overline{D})$ should be close to unity in the case of low decision freedom, whereas it should exceed unity under high freedom of choice.

Perceived behavioral desirability. According to Jones and Davis (1965), given the assumption that the actor intended to produce the observed behavior, the lower its apparent attractiveness or desirability, the more likely it is that a dispositional attribution will be made. For example, celebrating a religious holiday is usually considered more pleasant or desirable than driving to church in a snowstorm. The desirability hypothesis suggests that an actor will be seen as more religious when he has been observed driving to church in a snowstorm than when he has been observed celebrating a religious holiday. From a Bayesian point of view, a desirable behavior is likely to be performed whether or not the actor has the disposition in question [that is, $p(B|D) \approx p(B|\overline{D})$]. In contrast, an unattractive behavior is more likely to be performed by a person having the appropriate disposition than by a person who does not [that is, $p(B|D) > p(B|\overline{D})$]. It follows that an undesirable behavior will tend to be more diagnostic than a desirable behavior.

---

40. Contrary to expectations, however, the attributed positions themselves did not differ in the two conditions. This demonstrates that a probabilistic measure of belief strength may show different results than will a measure of belief position.

In an experimental test of the desirability hypothesis, Jones, Davis, and Gergen (1961) had subjects listen to a tape recording of a role-playing situation in which an actor was applying to become either an astronaut or a submariner. The instructions given to the actor (which were also recorded on the tape) made it clear that in the forthcoming job interview it would be desirable for the actor to respond in an inner-directed fashion (in the astronaut condition) or in an other-directed fashion (in the submariner condition). In the interview, the actor was heard to respond either in an inner- or in an other-directed fashion. After listening to the interview, subjects were asked to rate the actor's "true" other- or inner-directedness.[41]

The manipulations of this study resulted in a 2×2 analysis of variance design, which is shown in Table 5.9—together with the obtained results (high scores indicate attributed other-directedness). Note that undesirable behaviors, i.e., behaviors inappropriate for the job sought, had greater impact on posterior probabilities than did desirable behaviors. The actor who behaved in an other-directed manner was more likely to be viewed as truly other-directed when he applied to become an astronaut (for which his behavior was inappropriate) than when he applied to become a submariner. Similarly, for the actor who behaved in an inner-directed fashion, attributions of inner-directedness were more likely in the submariner than in the astronaut condition.

Table 5.9 Perception of Actor's Other- or Inner-Directedness (Adapted from Jones, Davis, and Gergen, 1961)

Interview condition	Desirable behavior	Actor's behavior	
		Other-directed	Inner-directed
Astronaut	Inner-directed	91.12	69.15
Submariner	Other-directed	69.26	43.92

Direct support for a Bayesian interpretation of these results has been provided by Trope and Burnstein (1973). These investigators used two job-interview situations in a manner similar to that of the Jones, Davis, and Gergen study. In one condition, for example, subjects received information about two applicants for a job "teaching current affairs in a Jewish Sunday School in Detroit." One applicant expressed a pro-Israeli position in the job interview (desirable behavior), the other an anti-Israeli position (undesirable behavior). As in the Jones, Davis, and Gergen study, subjects revised their estimates of the applicant's true position to a greater extent when the applicant performed the undesirable be-

41. It appears reasonable to assume that in this study prior probabilities that the actor was inner- or other-directed did not differ across conditions, and hence the posterior probabilities can serve as an index of attribution.

havior. Further, Trope and Burnstein obtained measures of the likelihood ratios for desirable and undesirable behaviors by asking subjects to indicate the probability that a person would take the position advocated by the job applicant if he had a pro-Israeli attitude [$p(B|D)$] and if he had an anti-Israeli attitude [$p(B|\overline{D})$]. As expected, the subjective likelihood ratio $p(B|D)/p(B|\overline{D})$ was higher for the undesirable than for the desirable behavior.

Freedom versus desirability. Jones and Harris (1967) attempted to examine the simultaneous effects of perceived decision freedom and behavioral desirability. Subjects read a short essay on Castro's Cuba which was either pro-Castro (with assumed low desirability) or anti-Castro (high desirability), and its author (the actor) was alleged to have written it either under conditions of free choice (high decision freedom) or by assignment of a course instructor (low decision freedom). On the basis of this essay, subjects were asked to infer the actor's attitude toward Castro's Cuba. Unfortunately, this posterior estimate that the actor is in favor of (or against) Castro's Cuba cannot be directly used as a measure of attribution. Although Jones and Harris recognized that attribution must be defined as information *gained,* i.e., as a revision in beliefs, no estimates of prior probabilities were obtained. Without a prior-probability estimate, it is not clear whether the posterior attribution of a pro-Castro attitude represents a greater, the same, or a smaller revision in subjective probability than attribution of an anti-Castro attitude.[42] To obtain a rough measure of revision, it is possible to use the subject's own attitude toward Castro's Cuba as an estimate of the actor's perceived attitude prior to obtaining information about his behavior. On the average, the subjects' attitudes were somewhat negative toward Castro's Cuba (a score of approximately 32 on a scale ranging from 0 to 70). Estimates of revisions are obtained by subtracting this score from the attributed posterior attitudes in the different experimental conditions.

The results indicate that perceived attitudes were revised in the direction of the position adopted by the actor in writing his essay. The anti-Castro essay led to downward revisions, the pro-Castro essay to upward revisions. Interestingly, even when the actor was described as having had no freedom of choice, subjects tended to attribute his behavior to his attitude. The greatest revision, however, occurred in the condition where the actor performed an undesirable behavior under high decision freedom. Revisions in the remaining experimental conditions were lower and approximately equal.

Prior behavior probabilities. In an interesting monograph, Steiner (1970) linked perceived decision freedom to the perceived desirability of behavioral alternatives. According to Steiner, a person will be perceived to have had high decision freedom to the extent that his available options were equally desirable.

---

42. An even more serious confounding effect due to neglect of prior probabilities can be found in a study by Jones, Worchel, Goethals, and Grumet (1971) dealing with similar issues.

It can also be argued that a behavior's perceived desirability is related to its perceived probability of occurrence. That is to say, the probability that an actor will perform a given behavior will tend to increase with the behavior's desirability. As with desirability, therefore, a negative relation between prior probability and dispositional attribution may be expected. That is, dispositional attribution will be made with greater confidence as the behavior's prior probability decreases (Jones and Gerard, 1967; Jones and Harris, 1967, Ajzen, 1971a).[43]

To test these notions, Ajzen (1971a) manipulated desirability of chosen alternatives and perceived decision freedom in four hypothetical situations. Under high decision freedom, the actor had to make a choice from three desirable and one undesirable alternative; under low decision freedom, his choice was from three undesirable and one desirable alternative. Subjects were told that the actor had chosen either a desirable or an undesirable option.

For example, in one situation the actor could perform one of four behaviors on a Friday night.

High decision freedom		*Low decision freedom*	
1. Have dinner with a date	(+)	1. Have dinner with a date	(+)
2. Go to a party	(+)	2. Read assignments for next week's classes	(−)
3. Listen to some new record albums at a friend's place	(+)	3. Attend an optional evening quiz section	(−)
4. Inspect a friend's stamp collection	(−)	4. Inspect a friend's stamp collection	(−)

In both conditions of decision freedom, half the subjects were told the actor had chosen alternative 1 (desirable), the other half that he had chosen alternative 4 (undesirable).

Subjects indicated the probability that the actor had a particular disposition (e.g., that he enjoys eating out) after his choice was disclosed [$p(D|B)$]. Further, measures were obtained of $p(B)$—the prior probability of the behavior; of $p(D)$—the prior probability of the disposition; and of $p(B|D)$. Consistent with expectations, the chosen behavior's desirability as well as perceived decision freedom strongly affected the strength of dispositional attributions. Revisions from prior to posterior probabilities, that is, $p(D|B) - p(D)$, increased with perceived decision freedom and decreased with the chosen alternative's desirability.

Ajzen also found that desirable behaviors were perceived to have higher probabilities of occurrence than did undesirable behaviors. Further, an average correlation of $-.62$ across the four situations supported the prediction of a negative

43. A negative relation between prior probability of an event (p) and its informational value (h) or diagnosticity is also postulated in information theory where $h = -\log_2 p$.

relation between a behavior's prior probability and its diagnostic value. That is, the higher the behavior's prior probability, the lower was the revision in attribution of a disposition. Finally, posterior probabilities predicted on the basis of Bayes's theorem were found to correlate significantly with obtained posterior-probability judgments. The average correlation over the four situations was .52.

Multiple Dispositions

So far we have considered the conditions under which an observed behavior will be attributed to a given disposition of the actor, rather than to some external factor. Frequently, however, it may be possible to invoke more than one disposition to account for the behavior. The analysis presented by Jones and Davis (1965) assumes that any outcome or effect produced by the behavior can provide the basis for a dispositional attribution. One important factor is the extent to which the action leads to *unique* outcomes, i.e., to outcomes that would not be produced by alternative behaviors. The greater the extent to which a given effect is uniquely associated with the actor's chosen alternative, the more likely it is that a dispositional attribution will be made on the basis of that effect.[44]

Suppose that a person is planning a trip to Europe and that there are four package tours available within an acceptable price range. Each tour visits four major European cities. Assume that the actor is observed to select Tour A, which includes Berlin. According to the uniqueness principle, if Berlin is a unique outcome of Tour A, an observer would be more likely to infer that the actor is interested in visiting Berlin than if all four tours visited Berlin.

In terms of a Bayesian analysis, uniqueness of outcomes should affect the diagnostic value of the behavior. Let D stand for the actor's disposition (e.g., his desire to visit Berlin) and B for his behavior (e.g., his choice of Tour A). Now consider the situation in which Berlin is uniquely associated with Tour A. If the actor desires to visit Berlin, the probability that he will choose Tour A, that is, $p(B|D)$, should be relatively high. Conversely, if he does not desire to visit Berlin, the probability of his choosing Tour A, that is, $p(B|\overline{D})$, should be relatively low. The likelihood ratio $p(B|D)/p(B|\overline{D})$ would in this case be greatly above unity. In contrast, when Berlin is a common outcome of all tours, the probability that the actor will select Tour A should be the same, whether he desires to visit Berlin or not. In this case, the likelihood ratio will be close to unity. It follows that a behavior with a unique outcome is more diagnostic than one with an outcome common to other behavioral alternatives, and Bayes's theorem would predict stronger dispositional attribution in the former case.

---

44. The confidence with which such an attribution will be made should increase as the number of effects uniquely associated with the chosen alternative decreases (see Chapter 2).

To test this hypothesis, Ajzen [45] constructed four hypothetical situations similar to the touring example discussed above. In each situation, an actor was said to have chosen one of four alternatives. To manipulate uniqueness, Ajzen created four conditions such that one outcome was unique to the chosen alternative in Condition 1 and appeared in two, three, or all four alternatives in Conditions 2, 3, and 4. A condition with intermediate uniqueness is shown in Table 5.10 for a situation dealing with choice of a political candidate. The other three situations involved choice of a tour (described on p. 200), choice among four pro-

Table 5.10 A Condition of Intermediate Uniqueness in Ajzen's Study

Four candidates compete for a seat in the United States Senate. Mr. K must decide who to vote for. In their campaigns the candidates have expressed their positions on several issues, as follows:

Candidate A	*Candidate B*
Favors universal free medical services	Favors legalization of marijuana
Favors a guaranteed annual income	Opposes wage and price controls
Favors immediate troop withdrawal from Vietnam	Favors universal free medical services
Opposes wage and price controls	Favors immediate troop withdrawal from Vietnam
Candidate C	*Candidate D*
Favors universal free medical services	Favors immediate troop withdrawal from Vietnam
Favors immediate troop withdrawal from Vietnam	→ Opposes busing of school children
Favors increased military spending	Favors guaranteed annual income
→ Opposes busing of school children	Favors legalization of marijuana

Arrow signifies crucial aspect.

grams of study, and choice of an accurate personality description.

For each situation, subjects were asked to estimate $p(D|B)$, the probability that the actor had a particular disposition (e.g., that he was against busing of school children or that he was interested in visiting Berlin), given that he had chosen a certain alternative. In addition, measures were obtained of $p(D)$, $p(B|D)$, and $p(B|\overline{D})$. The results provided support both for the uniqueness hypothesis and for a Bayesian interpretation of the attribution process. Consistent with expectations, revisions in beliefs were found to increase with the outcome's uniqueness. In a similar fashion, the likelihood ratio $p(B|D)/p(B|\overline{D})$ also increased significantly with the extent of uniqueness.

45. Unpublished study. See also Ajzen and Holmes (1974).

Attribution to Self and Other

So far we have not made a distinction between attributions to another person as opposed to self-attributions. Since an actor may serve as an observer of his own behavior, an attributional analysis appears appropriate in both instances.

Beginning with Schachter's (1964) work on the determinants of emotional states, investigators have applied the notion of self-attribution to such areas as obesity (Schachter, 1971), insomnia (Storms and Nisbett, 1970), and psychotherapy (Valins and Nisbett, 1971). For example, Storms and Nisbett asked insomnia volunteers to take a pill on two consecutive nights. In one condition the pill was described as a drug that would *increase* arousal whereas in a second condition it was described as a drug that would *decrease* arousal. (Actually the pill was a placebo with no physiological effects.) The assumption was that arousal is one reason for the insomniac's difficulty in falling asleep. In the first experimental condition, subjects could now attribute their arousal to an external factor (the pill), but in the second condition they had to attribute it to themselves. As expected, after taking the pill, subjects who could attribute their arousal to an external factor took less time to fall asleep than did subjects who attributed the arousal to themselves.

Thus there is evidence that people can attribute dispositional states to themselves on the basis of their own behaviors. However, disagreement exists as to the extent to which processes of self-attribution resemble those of attribution to another person.

The primary controversy revolves around the amount and kind of information available to actor versus observer and around the relative importance placed on different items of information. Bem (1965, 1967, 1968b) has argued that the behavior and its effects constitute, in most cases, the major cues for making an attribution to the actor. Since these cues are available to both the observer and the actor, the two should make the same attributions. In contrast, others (R. A. Jones *et al.*, 1968; E. E. Jones and Nisbett, 1971) have proposed that different information is available to actor and observer, and even when they have the same information, they tend to focus on different aspects of the behavioral situation. Specifically, Jones and Nisbett argued for a unidirectional difference such that "there is a pervasive tendency for actors to attribute their actions to situational requirements, whereas observers tend to attribute the same actions to stable personal dispositions" (1971, p. 2). According to Jones and Nisbett, this difference will obtain not only because the actor has more detailed knowledge of the circumstances but also because different aspects of the available information are salient for actor and observer. Specifically, the actor's behavior is assumed to be most salient for an observer, whereas situational factors take on greater salience for the actor himself.

A Bayesian analysis also leads to the expectation that actor and observer may sometimes differ in their attributions. From a Bayesian standpoint it is not unreasonable to expect that actor and observer will sometimes differ in their

perceptions of a behavior's diagnosticity. As a result, they may not revise their beliefs about the presence of certain causal factors to the same extent. For example, when an actor is aware that he has little freedom of choice, his behavior has low diagnostic value for a dispositional self-attribution. However, an observer may be unaware of the actor's restricted freedom and may thus view the same behavior as highly diagnostic. In other situations the opposite may be true, such that the actor will view his own behavior as more diagnostic for a dispositional attribution than will the observer.

Several investigations have found significant differences between judgments of own versus other's ability following success or failure on a task (E. E. Jones *et al.*, 1968; Feather and Simon, 1971a; Jones and Nisbett, 1971). Frieze and Weiner (1971), however, have reported no significant differences between actor and observer in terms of causal attributions for successful or unsuccessful task performance. More important, Feather and Simon (1971a) showed that differences between actor and observer depend on the circumstances. Consistent with the Jones-Nisbett hypothesis, subjects were found to attribute another person's success more than their own success to ability. However, contrary to that hypothesis, another person's failure was judged to be due to external factors (such as bad luck) more so than was own failure.[46]

Bem's (1965, 1967) findings using the interpersonal replication procedure described earlier also appear to be inconsistent with the hypothesis advanced by Jones and Nisbett. In a number of studies Bem found that his observers were able to infer with considerable accuracy the original subject's responses in a dissonance experiment. He thus concluded that results obtained in dissonance experiments could be explained in terms of the subject's self-attribution without recourse to assumptions about a motivational state such as dissonance.[47] It may therefore be argued that actors and observers appear to make very similar causal attributions.

One difficulty with this argument is the fact that Bem (1965, 1967) dealt only with posterior-probability estimates. In terms of posterior probabilities, differences may obviously be expected whenever actor and observer hold different prior probabilities or beliefs. In fact, because of the actor's greater knowledge of his own past behaviors, feelings, and experiences, actor and observer are quite likely to hold different prior beliefs about the actor's dispositions. For example, after performing the experimental task (but prior to the counterattitudinal behavior), subjects in the Festinger and Carlsmith (1959) study described earlier tended to believe that the task was somewhat boring. An observer, however, may

46. The means in the direction of external attribution for success were 5.5 (self) and 4.8 (other); for failure 4.1 (self) and 5.8 (other). Unfortunately, no direct test of the relevant interaction was reported.

47. Other investigators, however, have not always been able to replicate Bem's findings (Dillehay and Clayton, 1970; Piliavin *et al.*, 1969; R. A. Jones *et al.*, 1968). A Bayesian analysis of dissonance research will be found in Chapter 10.

at this stage assume that the actor believed the task to be quite interesting. It is thus not clear that the posterior estimates made by actor and observer after performance of the counterattitudinal behavior represent the same amount of *revision* in beliefs, i.e., that they reflect the same attributions. Thus the subject's judgment of the task as quite interesting in the $1 condition of the Festinger and Carlsmith study may represent a considerable revision from his prior belief that it was somewhat boring. An observer's posterior belief that the actor found the task quite interesting, however, may reflect relatively little change in his judgment.

Dissonance theorists have criticized Bem's analysis precisely for this reason (e.g., R. A. Jones *et al.*, 1968). To demonstrate the importance of prior beliefs, observers were given information about the original subject's initial (i.e., prior) judgment. Not surprisingly, this information influenced the observers' inferences, and more important, the observers could no longer infer the actual subject's responses.

In an interesting refutation of these arguments, Bem (1968b) insisted that observers in the interpersonal replication procedure should *not* be given information about the subject's initial judgments since Bem assumed that the subject himself is not aware of his initial position at the time that he provides his posterior responses. In support of this claim, Bem and McConnell (1970) found that only few subjects in a dissonance experiment could accurately recall their initial beliefs.

To summarize, it appears that actors and observers may form different inferential beliefs on the basis of the actor's behavior. However, evidence to date does not indicate that different inference processes need to be invoked in order to explain the observed differences in inferential beliefs. Attributions to both self and other can be accounted for in a Bayesian framework. Observed differences in posterior beliefs can easily be explained in terms of differential availability of information affecting prior beliefs. Differences in causal attributions (or revisions in beliefs) will be found when information available to actor and observer lead to differences in the behavior's perceived diagnostic value.

Error and Bias in Attribution

The application of Bayes's theorem to causal attribution implies an observer who processes information about an actor's behavior in a rather rational and objective manner. Yet it should be obvious that his causal explanations do not always correspond to reality. For one thing, the information on which his inferences are based may be misleading. That inaccurate information will produce erroneous attributions was demonstrated by Valins (1966), whose subjects were given false feedback about changes in their heart rates while viewing slides of seminude females; pictures associated with heart-rate changes were judged to be more attractive.

Misperception of relevant information may occur without actual deception. We have noted a tendency for probability judgments to be conservative (W. Edwards, 1968) and to take insufficient account of sample size (Kahneman and Tver-

sky, 1972). In many attribution situations it is impossible to tell what the correct inference would have been. A case in point is the study by Jones and Harris (1967) mentioned earlier, in which subjects tended to infer attitudes from expressed positions under no-choice conditions. It may be argued that even when the actor is assigned a given position, he exercises a certain degree of freedom in accepting the assignment and in the way he executes it. Subjects may thus be quite correct when they infer that the actor's attitude corresponds to some extent to the position he has advocated.

The attributional errors discussed thus far follow consistently from the observer's reasonable, albeit incorrect, beliefs about the situation. His misjudgment of a given behavior's diagnosticity may be construed as an "honest mistake." By way of contrast, attribution theorists have sometimes postulated certain more irrational tendencies on the part of the observer. The general idea is that a person's needs, values, and desires may influence his attributions.

Maintaining a Positive Self-Image

One kind of bias assumes a desire on the part of the observer to maintain a positive self-image. According to this view, ego-defensive biases will become operative whenever explanation of an observed event has positive or negative implications for the observer.

A study by T. J. Johnson, Feigenbaum, and Weiby (1964) has sometimes been taken to demonstrate defensive attribution. Subjects served as arithmetic teachers on two consecutive trials. The simulated student did poorly on the first trial. On the second trial the student continued to perform poorly for some teachers, and he improved for others. Teachers tended to attribute continued failure to the student whereas they attributed improved performance to themselves. Although these findings are consistent with a tendency toward ego-defensive attribution, alternative interpretations are available. For example, Kelley (1967) suggested that insofar as the teacher made a special effort on the second trial, the covariation of effort and success in the improvement condition would favor internal attribution, and lack of such covariation in the continued failure condition would favor external attribution. Alternatively, Kelley's notion of consistency suggests that consistent behavior across occasions is attributed to some factor internal to the actor whereas inconsistent behavior is attributed to an external factor. It follows that the student's consistent failure on the task should be attributed to an internal factor, such as his lack of ability. When he first fails and then succeeds (inconsistent performance), an external attribution to the teacher is likely. As noted earlier, these explanations are consistent with a Bayesian analysis, and they suggest a rational rather than defensive attribution process.

A related finding reported by Streufert and Streufert (1969) is that subjects tend to attribute their own success on a task to internal factors and failure to external factors. At first glance this finding also appears to indicate a defensive inference process in that subjects seem reluctant to accept blame for failure while

accepting praise for success. It should be clear, however, that the hypothesis of defensive attribution is applicable only in the case of self-attribution. The finding that the same difference in causal attribution obtains when subjects are asked to account for some other person's performance (Frieze and Weiner, 1971) must therefore be interpreted as inconsistent with the notion of ego-defensive attribution.

If an ego-defensive bias appears untenable, Bayes's theorem suggests two alternative explanations of the obtained results. The first assumes that subjects perceive the experimental task as relatively difficult, with the effect that success has greater diagnostic value for the actor's ability than failure has for his lack of ability. As a result, success would be attributed to ability more than would failure to lack of ability.[48]

The second explanation would hold if subjects expected to succeed on the task. Feather (1969; Feather and Simon, 1971b) has either measured or manipulated the subject's prior expectation that he would succeed or fail on an anagram task. Actual performance consistent with expectations was attributed to stable internal factors (skill and ability), but performance inconsistent with expectations was attributed to unstable external factors (luck). Thus, when subjects expected success, they displayed the tendency found by Streufert and Streufert to attribute actual success to their own ability more than actual failure to their lack of ability. However, contrary to an ego-defensive bias but consistent with a Bayesian analysis, when subjects expected to fail, they attributed actual failure to their own lack of ability more than they attributed actual success to their ability.

Bayes's theorem thus accounts for the observed difference between inferences based on success and failure in terms of information processing without reference to irrationality or defensiveness.

Hedonic Relevance

A second type of bias is assumed to be elicited by factors that increase the observer's need to explain the event he has witnessed. An actor's behavior may sometimes have important rewarding or punishing implications for the observer. According to Jones and Davis (1965), such "hedonic relevance" increases dispositional attribution; the observer will tend to attribute the behavior to some dispositional factor internal to the actor.

Jones and deCharms (1957) investigated the effects of hedonic relevance in two experiments. A confederate working with four or five subjects always failed on a series of problem-solving tasks. In the low-relevance condition, a $1 reward was dependent simply on the subject's own success whereas in the high-relevance

48. Let S stand for success, F for failure, and A for ability. Given a difficult task, this argument assumes the following relation between the likelihood ratios:

$$p(S|A) \,/\, p(S|\bar{A}) > p(F|\bar{A}) \,/\, p(F|A).$$

condition, the reward was given only when all group members succeeded. The results provided only partial support for the predicted effects of hedonic relevance on dispositional attribution. In the first experiment, hedonic relevance decreased the perceived competence and dependability of the confederate, but it did not affect his perceived motivation. The second experiment provided even less support in that the predicted effect was significant only for attribution of dependability. This investigation therefore provides little support for motivational biases in attribution.

Attribution of Responsibility

A different motivational bias was postulated by Walster (1966) with respect to attribution of responsibility for an accident. According to Walster, a person's need to assign responsibility for an accident to the actor who has caused it increases with the severity of its consequences. Using descriptions of a hypothetical car accident, she manipulated severity of damage to a car and injuries to others.

Subjects listened to one of four tape recordings describing Lennie, a high school student, whose car rolled off a hill because of a brake failure. Walster manipulated mild and severe consequences in four experimental conditions: (1) The car's fender was only mildly damaged, but the car could have been completely destroyed; (2) the car was completely destroyed; (3) the car's fender was only mildly damaged, but bystanders could have been severely injured; and (4) bystanders were severely injured.

Consistent with predictions, the finding was that more responsibility was attributed to Lennie in Conditions 2 and 4 (severe consequences) than in Conditions 1 and 3 (mild consequences). The results, however, were not unequivocal; the difference between Conditions 3 and 4 only approached significance for the total sample and was not significant for female subjects. Further, it was found that injury to bystanders (severe consequence) did *not* lead to more attribution of responsibility than damage to the car (mild consequence). Most important, in subsequent studies, Walster (1967) as well as other investigators (Shaver, 1970a, 1970b; J. I. Shaw and Skolnik, 1971) failed to replicate the effect of severity on attribution of responsibility.

Perhaps because of the obvious implications for the judicial process, investigators have shown considerable interest in factors influencing attribution of responsibility. Much of the research in this area has been based less on systematic theoretical analyses of attribution processes than on intuitive hypotheses and speculations. Perhaps for this reason, studies on attribution of responsibility have yielded contradictory and inconclusive results.

According to the analysis by Jones and Davis (1965) discussed earlier, the severity or desirability of an act's outcomes will result in dispositional attribution only when the observer assumes intentionality on the part of the actor. Since the situations used in studies on attribution of responsibility for an accident clearly imply lack of intentionality on the part of the actor, Jones and Davis's analysis

suggests that variations in outcome severity should *not* influence attribution of responsibility.

A number of other variables that might influence attribution of responsibility have been investigated: attractiveness of actor and victim (Landy and Aronson, 1969; Shepherd and Bagley, 1970), similarity of actor and subject (Shaver, 1970a; Shaw and Skolnik, 1971), whether or not the actor had insurance (Shaver, 1970b), etc. Generally speaking, results of these studies have been highly inconsistent and have not led to any firm conclusions. Partly in an attempt to understand these conflicting findings, investigators have obtained additional measures designed to assess the subject's perceptions of the hypothetical situations. Thus subjects have been asked whether the actor had been careless, whether he was conscientious, whether he could have foreseen the consequences, etc. Unfortunately, these questions have shed little light on attributions of responsibility.

A number of points need to be made with respect to these investigations. First, the term "attribution of responsibility" is somewhat misleading since responsibility can be described neither as a disposition of a person nor as a property of an object. Attribution of responsibility can perhaps best be viewed as a moral judgment. Moreover, Heider (1958) has made it clear that the term "responsibility" can be interpreted in different ways. As mentioned in Chapter 2, Heider distinguished five levels of responsibility that correspond in part to Piaget's (1932) developmental stages in moral judgment.

1. *Association.*[49] At the first and most primitive level, the actor is held responsible for any effects that are in any way associated with him.
2. *Commission.* At the next level, the actor is held responsible if he was instrumental in producing the observed effects (even if he could not have foreseen them).
3. *Foreseeability.* At this level, the actor is held responsible only if he could have foreseen the effects even though he might not have intended to produce them.
4. *Intentionality.* At the fourth level, the actor is held responsible only for effects he foresaw and intended.
5. *Justification.* Finally, at the fifth level, he is held responsible only to the extent that his intended behavior was not justifiable, i.e., not caused by factors beyond the actor's control.

At each successive level, attribution to the person becomes less likely, attribution to the environment more likely. One immediate implication of Heider's analysis is that general measures of responsibility will tend to be ambiguous. A question such as "Is the actor responsible for the accident?" could be interpreted in different ways: (1) Was the actor associated with the accident? (2) Was he

49. The labels used here for the five levels of responsibility were introduced by Shaw and Sulzer (1964).

instrumental in producing the accident, i.e., did he cause it? (3) Is he responsible in the sense that he could have foreseen the accident? (4) Did he intend to cause the accident? (5) To what extent was his behavior justified? Since subjects are usually not told at what level they are to respond, judgments of responsibility may take on different meanings in different conditions of an experiment, in different investigations, and for different subjects.

Shaw and his associates (Shaw and Sulzer, 1964; Shaw and Reitan, 1969) have used Heider's levels of responsibility in a somewhat different way. Their analysis suggests that not only can a *subject's responses* vary along these five levels but so can the *behavioral context.* That is, the context itself may suggest that the actor was merely associated with the event, that he committed (caused) the event, etc. If the fourth and fifth contextual levels are reversed, a two-way classification scheme can be constructed; [50] this scheme is presented in Table 5.11.

Table 5.11 Attributions of Responsibility in Terms of Heider's Levels of Causality

	Contextual level				
Developmental or response level	Association	Commission	Foreseeability	Justification	Intentionality
Association	X	X	X	X	X
Commission		X	X	X	X
Foreseeability			X	X	X
Intentionality				X	X
Justification					X

An X indicates that a judgment of responsibility would be made.

The X's in Table 5.11 indicate the combinations of developmental and contextual levels that should lead to attributions of responsibility. For example, a subject at the development level of commission (or a subject who believed that he was to interpret responsibility in terms of commission or instrumentality) would not attribute responsibility when the context merely associated the actor and the event, but he would attribute responsibility at the remaining contextual levels. In contrast, a subject at the fourth developmental level (intentionality) would attribute responsibility if the context suggested that the actor behaved intentionally, whether his behavior was justifiable or not. At the highest developmental level of justification, attribution of responsibility would occur only when the context suggested that the actor behaved intentionally without justification. It can thus be seen that the proposed classification scheme represents the pattern of a perfect

50. Actually, it is not clear that the contextual level of justification should be placed between foreseeability and intentionality. It is possible that justification leads to a level of attribution that stands between commission and foreseeability. Shaw and Reitan (1969) reported data that appear to support the latter order.

Guttman scale, suggesting that Heider's levels of causality form a single dimension of perceived responsibility.

This analysis implies that, unlike attributions of dispositions, attributions of responsibility can be made even when the actor is not perceived to have behaved intentionally. For example, at the developmental or response level of commission, an actor will be held responsible if he was instrumental in bringing about the observed effect, even though he may not have intended to produce it. More important, our analysis makes it clear that attribution of responsibility is a function of two factors: the observer's developmental or response level and the contextual level. When the level of one factor is held constant, attribution of responsibility should depend on the level of the other factor.

We can illustrate this principle in terms of the judicial system, where response level is determined by law and assignment of responsibility or guilt is determined by the contextual level. Thus, in a murder trial, the judge will carefully instruct the jury as to the response level they are to adopt. A guilty verdict requires intentionality when the charge is murder but not when the charge is manslaughter. Further, the law allows for justified homicide under certain circumstances. In an accident case, instrumentality or commission may be the response level specified by law for a judgment that the actor is guilty.

In these situations, then, the response level is clearly specified, and the observer must judge the contextual level in order to attribute responsibility. Alternatively, it is possible to specify the contextual level and let the response level be determined by the observer. This procedure was followed by Shaw and his associates (Shaw and Sulzer, 1964; Shaw and Reitan, 1969). Shaw and Sulzer used 20 short behavioral situations varying along the five contextual levels. For example, one description at the level of association read as follows: "A boy hit another child with Perry's toy gun. Is Perry responsible for the child being hit?" A description at the intentional level stated: "Perry opened the door so that the wind would blow the children's papers all over the room. Is Perry responsible for the scattered papers?" Attributions of responsibility were made on a six-point scale.

On the assumption that subjects work at a *given* developmental or response level, attribution of responsibility should increase with the contextual level (see Table 5.11). The results of two experiments support this prediction (Shaw and Sulzer, 1964; Shaw and Reitan, 1969). Further, for subjects working at *different* developmental or response levels, different profiles of responsibility assignment should be obtained. In support of the classification scheme shown in Table 5.11, Shaw and Sulzer (1964) reported different response profiles for second-grade children than for college students. The former appeared to work at the level of commission, the latter at the level of intentionality. Evidence that adults work at the level of justification was found by Shaw and Reitan (1969). Second-grade children were found to assign responsibility in situations that involved at least commission, and as expected, their attributions did not increase at the higher contextual levels. On the other hand, adults tended to attribute more responsi-

bility at each successive contextual level. These findings are summarized in Fig. 5.6.

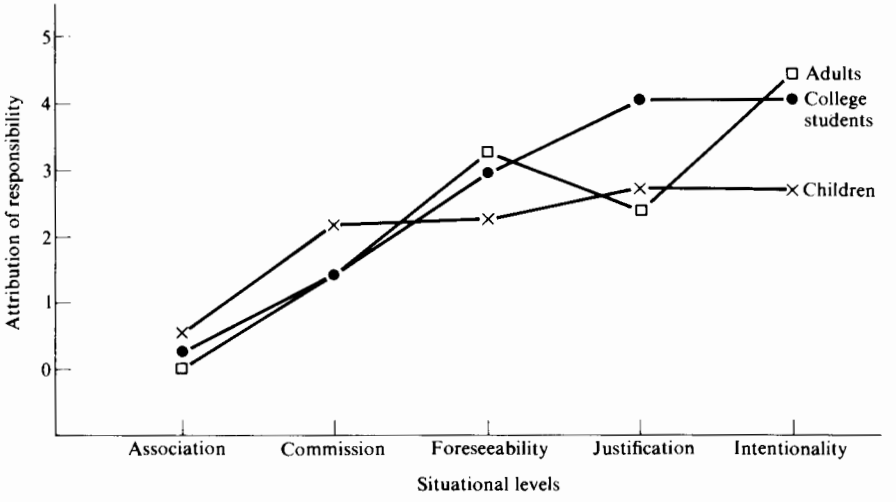

Fig. 5.6 Attribution of responsibility as a function of situational levels. (Adapted from Shaw and Sulzer (1964) and Shaw and Reitan, 1969.)

Shaw and Sulzer also varied outcome desirability (positive or negative), and consistent with expectations based on attribution theories, outcome desirability had its largest effect on attribution of responsibility when the action was intentional and a somewhat weaker effect when the outcome was foreseeable. At the other three contextual levels, outcome desirability had no effects on attribution of responsibility. These results are shown in Fig. 5.7. Similar results were found by Shaw and Reitan (1969), who varied not only outcome desirability but also intensity of outcomes (mild versus severe). The latter variable interacted with contextual levels, indicating that effects of severity on attribution of responsibility are greatest at the level of foreseeability.

As is apparent, when contextual levels are specified, assignment of responsibility follows predictable patterns. Further, as the comparison between children and adults shows, when developmental or response levels are specified, attribution of responsibility will also be made in a meaningful way. The conflicting and inconclusive results of most studies dealing with attribution of responsibility can now perhaps be appreciated since these studies have specified neither response levels nor contextual levels. For example, several studies have used Walster's (1966) description of an accident in which a car rolled down a hill. In this situation it is made quite clear that the effects were not *intended;* however, the contextual level may still involve either foreseeability, commission (instrumentality), or association. Furthermore, since subjects are given no instructions concerning response level, any of the five response levels may be involved. Judg-

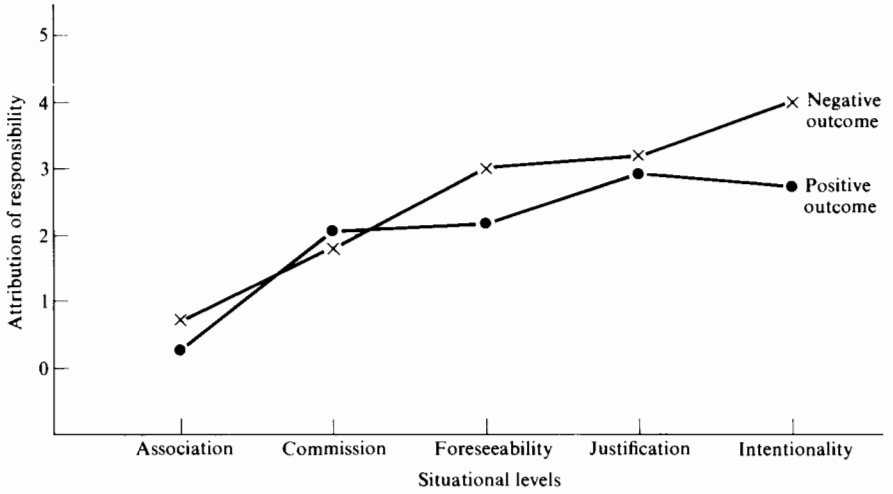

Fig. 5.7 Attribution of responsibility for acts with desirable and undesirable outcomes. (Adapted from Shaw and Sulzer, 1964.)

ments of responsibility in these studies are obviously uninterpretable, and conflicting findings are hardly surprising.

Summary of Attribution Research

Attribution theory deals with the ways in which a person arrives at an explanation for observed behavior. This attribution process is accompanied by and corresponds to revisions in the observer's beliefs about the actor and his environment. We showed that causal attribution and corresponding revisions in beliefs can be described in terms of Bayes's theorem. As one might expect from a Bayesian analysis, the observer's inferences generally tend to be quite realistic and to follow consistently from his prior beliefs about the situation. Although attribution theorists have often postulated certain more irrational dynamic processes, few studies have demonstrated these biasing effects convincingly.

Just as the observer makes causal attributions to explain the behavior of an actor, he can make self-attributions on the basis of his own behavior. To be sure, actor and observer may differ in their perception of a given behavior's diagnosticity, and they may therefore arrive at diverging causal attributions. However, attribution to both self and others can be accounted for in a Bayesian framework, and evidence to date does not indicate that different inference processes need to be invoked in order to explain differences in attribution between actor and observer.

Various factors have been proposed to affect causal attribution, including consistency across occasions, objects, and actors, uniqueness of behavioral effects, perceived desirability and probability of the behavior, and perceived decision freedom. We have reviewed some of the available evidence and shown that each of these factors tends to influence attribution as predicted. The different variables

can all be viewed as influencing the likelihood ratio in Bayes's theorem; that is, the variables treated by attribution theories suggest some of the conditions under which an observed behavior will have high or low diagnostic value for causal inferences. Bayes's theorem can therefore serve as a framework that permits a unified and integrated approach to the study of attribution processes.

CONCLUSION

In this chapter we have sought to identify and review various areas of research relevant to belief formation. Many of these lines of research have not received a great deal of attention from investigators in the attitude area. Owing to the wealth of material associated with each of these areas and to our own limited acquaintance with some of them, we have of necessity not provided an in-depth analysis of each line of research. Our main purpose in reviewing the different approaches is to acquaint the reader with some of the available information and to demonstrate its relevance for an understanding of belief formation. We will now attempt to summarize some of the major conclusions we have drawn in our review of the literature.

First, a distinction must be made between *descriptive* and *inferential* beliefs. Relatively little information is currently available about the formation of descriptive beliefs. What little evidence there is suggests that descriptive beliefs are formed on the basis of direct observation and tend to be fairly veridical; there is little evidence that personal factors such as beliefs, desires, attitudes, or personality characteristics have any systematic effects on the formation of descriptive beliefs. This conclusion appears to hold for beliefs about statements made by other people (i.e., beliefs that "*S* said *O* is *X*") as well as beliefs about attributes of objects or actions (i.e., beliefs that "*O* is *X*").

The distinction between descriptive and inferential beliefs is in many ways analogous to the distinction between *acquisition* of new information and using or *processing* that information. The latter distinction is found in research on multiple cue learning, where subjects must first learn the ecological validity of cues and then learn to use these cues consistently in making their judgments. Similarly, research on Bayes's theorem has distinguished between misperception (incorrect beliefs about a datum's diagnosticity) and misaggregation (inappropriate processing of new information). In the area of concept formation, a similar distinction was made between learning a concept's attributes and learning the appropriate combination rules.

In contrast to descriptive beliefs, personal factors do play a major role in the formation of inferential beliefs. Generally speaking, on the basis of one or more items of information—i.e., beliefs about an object, person, or action—individuals are found to form new beliefs through some inference process. A question of considerable importance is the extent to which inferences are made along an *evaluative* or *probabilistic* dimension. Although much attitude research assumes that inferences are based on evaluative consistency (i.e., that individuals attempt

to achieve evaluative consistency), our review of the literature provides little support for this assumption and suggests instead that the inference process is based primarily on probabilistic relations among beliefs. Although it is true that evaluative and probabilistic consistency are frequently indistinguishable, whenever it is possible to separate these two dimensions, inferences tend to follow lines of probabilistic reasoning.

This conclusion follows most directly from research on trait inferences, but it is also supported by findings in the areas of cue utilization and syllogistic reasoning. Perhaps most impressive, however, is the finding that various quantitative probability models provide fairly accurate descriptions and predictions of inferential belief formation. Research concerned with these models again provides little evidence that evaluative consistency has any direct effects on inference processes.[51]

It appears, then, that individuals acquire relatively stable conceptions of probabilistic relationships among sets of beliefs. This probabilistic structure serves as the basis for inferential belief formation. A new item of information about a given object or event will thus have implications for many other beliefs about the object or event. The most direct demonstration of this process is given by research on trait inferences and "implicit theories of personality." Similarly, observation of a given behavioral event was shown to have probabilistic implications for causal attribution; observation of an actor's behavior takes on differential diagnostic value with respect to different attributions.

It is worth emphasizing that inferential belief formation involves *revision* in subjective probabilities rather than subjective probabilities at a given point in time. Although it may be of interest to study the beliefs of a person at a given moment, attempts to understand the effects of a new item of information on those beliefs, i.e., attempts to understand inference processes, demands consideration of change in the beliefs produced by the new information. Clearly, unless one can assume that prior beliefs are equivalent across different experimental conditions, measures of posterior beliefs are insufficient for an understanding of the inference process.

Despite the apparently overwhelming evidence in favor of a probabilistic rather than evaluative basis for inferential belief formation, a cautionary note is in order. The experimental situations in which inferences have been studied have typically not dealt with highly involving or emotional issues. In many experiments the subjects are intelligent and educated college students who are capable of adopting a problem-solving approach and thus may attempt to provide logical or

---

51. Although not reviewed in this chapter, some indirect support for the relative weakness of evaluative inferences comes from studies that have used balance theory to predict inferences about interpersonal relations (Wyer and Lyon, 1970) or learning of them (Lewit and Shanley, 1969; Press, Crockett, and Rosenkrantz, 1970; Rubin and Zajonc, 1969). These studies found only weak support for predictions based on notions of affective balance.

rational responses. These conditions will favor probabilistically oriented inference processes.

Although emotional or affective factors may play a more important role in inferences about involving issues, evidence to date does not support this view, and we take the position that the inference process is based primarily on probabilistic consistency. Thus man may be viewed as a fairly rational processor of information available to him. His descriptive and inferential beliefs are not capricious, nor are they systematically distorted by motivational or emotional biases.