Attribution Theory in Social Psychology[1]

HAROLD H. KELLEY

University of California, Los Angeles

In 1958, Fritz Heider, in his book *The Psychology of Interpersonal Relations*, described the processes by which the individual makes attributions about his world—attributions of causes, dispositions, and inherent properties. This book continued a line of analysis Heider had begun in his earlier writing on phenomenal causality and traced out its implications for many aspects of social perception. In the meantime, attribution concepts have begun to appear in many areas of social psychology, used by persons of highly varied theoretical backgrounds in their analysis of quite diverse phenomena. In some cases, the analysis is explicitly adopted from Heider, as in the excellent recent paper by Jones and Davis (1965) on the attribution process in person perception. In other cases, the relation to Heider's concepts is not recognized. It seems that, like Molière's character and the use of prose, many social psychologists have been using attribution theory all their lives without knowing it. My purpose in this paper is to highlight some of the central ideas contained in Heider's theory, to present them in a systematic way, and to show their relevance to developments in several central fields of contemporary social psychology. In many ways, this effort is similar to that of Jones and Davis. It is hoped that this sort of analysis and explication will advance the development of attribution theory as an explicit, hypothesis-generating, and research-provoking set of principles.

[1] While I assume full responsibility for this paper, I gratefully acknowledge the many fruitful conversations on these topics with Melvin Seeman, John Thibaut, and the students in three graduate seminars. The preparation of the paper was facilitated by Grant GS-1121X from the National Science Foundation.

Attribution theory concerns the process by which an individual interprets events "as being caused by particular parts of the relatively stable environment" (Heider, p. 297). Consideration of attribution theory is relevant for a symposium on motivation in several respects. The theory describes processes that operate *as if* the individual were motivated to attain a cognitive mastery of the causal structure of his environment. Indeed, Heider explicitly assumes that "we try to make sense out of the manifold of proximal stimuli . . ." (p. 296). And Jones and Davis state, "The perceiver seeks to find sufficient reason why the person acted and why the act took on a particular form" (p. 220, emphasis omitted). This broad motivational assumption makes little difference in the development and application of the theory, but it gives the theory a definite functionalistic flavor ("man grasps reality, and can predict and control it" [Heider, p. 79]) and affords whatever motivational basis might seem necessary to support the complex cognitive processes entailed in attribution.

More important for the student of motivation are the specific processes and their consequences. Attribution processes are assumed to instigate, under certain conditions, such activities as information-seeking, communication, and persuasion. Thus, attribution theory plays an important role in describing the motivational conditions for these significant classes of social behavior. Equally important is the relevance of attribution theory to the *perception* of motivation, both in others and in one's self. A major application of the theory concerns the processes by which the typical observer infers a person's motivations from his actions. Although attribution theory deals with how the naïve person makes these inferences, it can be studied with profit by the personality diagnostician for the parallels it contains to his scientific assumptions and procedures.

The Attribution Process

Attribution refers to the process of inferring or perceiving the dispositional properties of entities in the environment. These are the stable features of distal objects such as color, size, shape, intention, desire, sentiment, and ability. Any given stable feature may manifest itself in many ways. A given rectangular surface

produces many specific retinal images. A person's desire to benefit me is expressed by many different actions. In our perception, we tend to interpret, analyze, and order this "variable manifold of mediating events" in order to achieve an understanding of the "contents of the distal environment" (Heider, p. 296).

This understanding is gained by way of a causal analysis that is "in a way analogous to experimental methods" (Heider, p. 297) and that has the purpose of disentangling which effects are to be attributed to which of several factors present. In the basic case, where the person is concerned with the dispositional properties of his surrounding environment, the choice is between external attribution and internal (self) attribution. This can be illustrated by an example concerning the attribution of object desirability *vs.* personal desire. Am I to take my enjoyment of a movie as a basis for an attribution to the movie (that it is intrinsically enjoyable) or for an attribution to myself (that I have a specific kind of desire relevant to movies)? The inference as to where to locate the dispositional properties responsible for the effect is made by interpreting the raw data (the enjoyment) in the context of subsidiary information from experiment-like variations of conditions. A naïve version of J. S. Mills' method of difference provides the basic analytic tool. The effect is attributed to that condition which is present when the effect is present and which is absent when the effect is absent. This basic notion of covariation of cause and effect is used to examine variations in effects (responses, sensations) in relation to variations over (a) entities (movies), (b) persons (other viewers of the movie), (c) time (the same person on repeated exposures), and (d) modalities of interaction with the entity (different ways of viewing the movie). The attribution to the external thing rather than to the self requires that I respond *differentially* to the thing, that I respond *consistently,* over time and over modality, and that I respond *in agreement* with a consensus of other persons' responses to it. In other words, the movie is judged enjoyable if I enjoy only it (or at least, not all movies), if I enjoy it even the second time, if I enjoy it on TV as well as at the drive-in theater, and if others also enjoy it. If these conditions are not met, there is indicated an attribution to the self (I enjoy all movies, or I

alone have a weakness for this particular type) or to some juxtaposition of circumstances (I was in an especially susceptible mood on the one occasion).

The logic of the analysis is obviously akin to that employed in analysis of variance and can be illustrated by the diagram in Figure 1. Along the vertical axis are listed *entities* which correspond to things in the environment (e.g., movies). Along one horizontal axis are various *persons* who interact with the entities and along the other, the various *modalities* of interaction and the *times* at which these interactions may occur. The X, Y, and

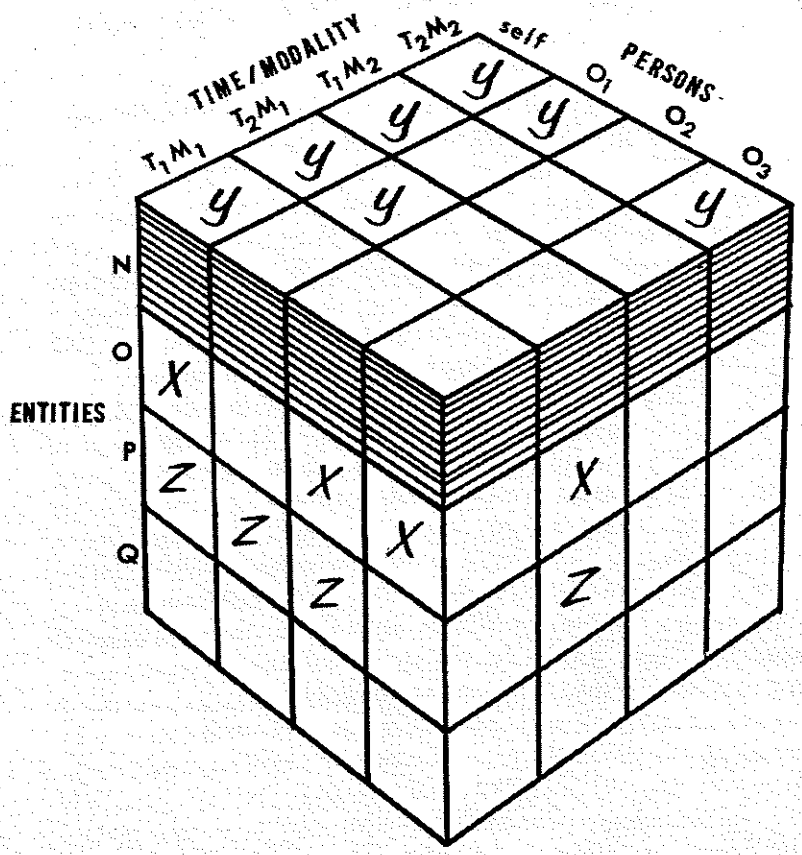

FIG. 1. Data pattern indicating attribution of effect Y to entity N.

Z's represent *effects* such as experiences, sensations, or responses. These effects are defined from a given person's point of view (the "self" in Figure 1) inasmuch as we are dealing with one person's assembly of information in order to make an attribution. This person makes the attribution of effect Y to entity N when, as in the figure, Y and Y alone occurs exclusively within the horizontal plane corresponding to N. This means that the effect Y varies only as between entities, is uniquely associated with N, and occurs consistently over time, modality, and persons when N is present. In general, we might say that the subjective criteria for the possession of valid knowledge about the external world are distinctiveness of response coupled with consistency and consensus.

Heider gives a number of similar examples which involve the assignment of an effect to a cause either in the environment or in the person. In addition to the problem of environmental *vs.* personal factors in enjoyment (in the illustration above), his examples include difficulty of task as an environmental factor *vs.* ability as a personal factor, objective requirements *vs.* personal wishes in relation to values and norms, and external *vs.* personal responsibility for a person's compliance to another's influence. In all these cases, the attribution is presumed to follow the rules outlined above: External attribution (problem difficulty; norms and values reflecting objective, invariant standards; external responsibility) is made when evidence exists as to the distinctiveness, consistency, and consensus of the appropriate effects.

I have emphasized the allocation of causality between the environment and the self because when cast in these terms the theory bears on the central issue in psychological epistemology; namely, the basis of subjective validity. But the analysis is also applicable to other cause-effect problems. For example, an observer may witness my pleasure at attending the movie and may consider whether to conclude something about me or about the movie. Or, I may have the problem of assigning responsibility for some effect, good or bad, between two other persons. Or the problem may be the one in social perception that Jones and Davis analyze, involving my inferring a person's intentions from knowledge of the consequences of his actions. Before considering at greater length some of these problems of social perception, attention needs to be directed to the implications of the basic self-environment analysis for a central area of social psychology, namely, social influence.

Informational Dependence and Influence

The attribution criteria are exceedingly useful in the analysis of informational dependence and persuasion. Here I wish briefly to outline an argument that John Thibaut and I have developed in our chapter on "Group Problem Solving" for a new edition of the *Handbook of Social Psychology*.

As noted above, in the analysis of external *vs.* internal ascription, attribution theory confronts the problem of the phenomenology of attribution validity. When a person has an impression that something is true of an entity, how does he ascertain that the impression reflects the inherent properties of the entity and not his own characteristics or some peculiar interaction with the entity? The four criteria for external validity are those mentioned above:

1. *Distinctiveness:* the impression is attributed to the thing if it uniquely occurs when the thing is present and does not occur in its absence.
2. *Consistency over time:* each time the thing is present, the individual's reaction must be the same or nearly so.
3. *Consistency over modality:* his reaction must be consistent even though his mode of interaction with the thing varies. (For example, he sees it to have an irregular outline and he feels it to be rough; or first he estimates the answer to the problem and then he calculates it.)
4. *Consensus:* attributes of external origin are experienced the same way by all observers.

To the degree a person's attributions fulfill these criteria, he feels confident that he has a true picture of his external world. He makes judgments quickly and with subjective confidence, and he takes action with speed and vigor. When his attributions do not satisfy the criteria, he is uncertain in his views and hesitant in action. It is not assumed that fulfillment of these criteria implies veridicality of the person's attributions. The specified evi-

dence provides a basis for subjective validity (as manifested, for example, in confidence in the validity of one's attributions) but not necessarily a basis for their objective validity.

With respect to the problem of information dependence, these concepts suggest an approach to indexing the individual's *state of information* regarding his world. The criteria suggest that this indexing be done in terms of (a) differentiation and (b) stability of attributions. Because Heider's criteria imply the examination of evidence derived from pseudo-experimental operations, we might expect to find parallels between an appropriate index of attribution validity on the one hand and statistical indexes used in the analysis of experimental data on the other. This is simply to highlight the analogy between the naïve analysis of evidence and the scientific one. The first criterion (distinctiveness) seems to correspond to the numerator or between-condition term in the usual *F* ratio and the last three criteria (consistency over time, modality, and persons) correspond to the error or within-condition term. As a measure, then, of a person's state of information regarding a given entity, the theory suggests an analogue of the *F* ratio in which the degree of differentiation between the various entities is compared with the stability of attribution (based on the consistencies and consensus) with respect to the given entity. In conceiving the denominator of this ratio, we are assuming that both the person's own consistency evidence and the evidence from social consensus contributes to the stability of his own reaction. The attribution he makes on any given occasion depends on some sampling of the information available to him, both from his own present and recent experience and from social sources. The more consistent this information is, the more stable will his attribution be. Insofar as this stability is high (or his variability in reaction is low), the "error term" in the ratio is small and, assuming he can make a more or less distinctive attribution to the entity, his information state is high. In brief, information level is high for a person who can make highly stable but differentiated attributions.

Information level, defined in this manner, provides a convenient concept in the analysis of information dependence. In our book *The Social Psychology of Groups*, Thibaut and I (1959), have made an analysis of dependence (and interdependence) in terms of outcomes (rewards and costs). That analysis assumes that we will ultimately have the capability meaningfully to measure or index each person's level of outcomes. We are now proposing that a similar analysis be made of information dependence (and interdependence). And we are similarly assuming that psychologists will ultimately provide a meaningful index of the person's level of information along the general lines suggested above.

Given this conception of information level, the basic statements in our analysis of information dependence are straightforward. Person A is informationally dependent upon B if B can raise A's level of information to a higher level than A can attain from alternate sources. As in the case of outcome dependence, information dependence can be defined *objectively*, in terms of potential or actual effects on A's information level of receiving communications from B, or *subjectively*, in terms of anticipated or experienced effects. Cutting across this distinction is one of time reference, "potential" and "anticipated" having a future reference and "actual" and "experienced" being oriented to the present or past.

Anticipated information dependence affords the basis for *seeking* information. Person A will seek information from B if he believes B's information can raise him to a level higher than can information from available alternative sources. Person A will particularly seek information from B if he believes the latter's messages will yield at least as high a level of information as he estimates to be possible for the particular type of entity or problem. In this regard, it seems plausible to assume that persons have conceptions of the level of information they may expect to achieve for various types of problems, tasks, or phenomena. For example, we expect high distinctiveness and consistency in the answers to arithmetic problems. If the answer to a homework problem my child shows me is not different from that to other problems and is not reproducible by different methods, by different persons, and on repeated occasions, I am highly doubtful of its validity. On the other hand, the outcomes of those tasks identified as games of chance are expected to be charac-

terized by low distinctiveness and high inconsistency. Given these different expectations regarding level of information attainable for a given type of task, we may expect persons to be dissatisfied with their information state when it falls below the expected level, and (assuming the task is of some importance to them) to initiate information-seeking activities. These activities will be directed toward a particular source (another person, a certain book, or a particular method of investigation) as long as the level of information believed to be attainable there is higher than that anticipated in interaction with alternative sources.

(The readers of Thibaut and Kelley will recognize in the above comments the possibilities for developing concepts in the area of information dependence that are parallel to the concepts in the outcome domain of *comparison level* or CL and *comparison level for alternatives* or CLalt.)

These concepts are useful in organizing a great deal of the social psychological literature on the conditions governing (a) susceptibility to persuasion, (b) immediate success of persuasion, and (c) the persistence of its effects.

Person A will be more susceptible to influence the more variable his prior attribution has been. Attribution instability (and, hence, susceptibility to influence) will be high for a person who has (a) little social support, (b) prior information that is poor or ambiguous, (c) problems difficult beyond his capabilities, (d) views that have been disconfirmed because of their inappropriateness or nonveridicality and (e) other experiences engendering low self-confidence. (The reader need not be reminded of the vast amount of social psychological research that bears upon these various points.)

With regard to B's success in persuading A, we may first assert that A will be influenced by B if B's message enables A subsequently to achieve a higher level of distinctiveness and stability in his attributions than before. However, the factors relevant to this influence and the persistence of its effects will depend upon which of two methods of influence B uses. As B tries to create stability for A for a new attribution, he may operate on either the *consistency* aspects or the *consensus* aspects of A's attribution

process. (Of course, he may do both, and persuaders usually do, but for simplicity I will consider only the two pure cases.)

In the first instance (usually called education or *instruction*), B attempts to provide A with the means of obtaining consistency, both over time and over modality, in the new attribution. For example, B may demonstrate a new way of looking at or interacting with the entity which enables A to find consistency in his subsequent confrontations with it. What B provides may consist of new analytic methods, problem-solving procedures, practice in the use of a given modality in order to increase its consistency, new perspectives and frameworks for evaluating items, training in discrimination and judgment, the suggestion of crucial comparisons in order to sharpen discriminations and evaluations, a demonstration of the relevance of facts and information that A had not previously appreciated, or instruction in relevant verbal labels. These methods have in common the property that they may be used by A independently of B (once A has learned them) and they contribute directly to the fulfillment of the consistency criteria and, in that manner, enhance the subjective validity of whatever attribution they lead to. The content of this influence will be adopted and held by A because it affords him attributional stability, and not because of B's expertness or credibility. (However, the latter may be relevant to A's willingness to pay attention to B at the outset.)

The second influence method (usually referred to as *persuasion*) is more thoroughly dependent on A's evaluation of B. In this case, B conveys information about his own conclusions, attributions, evaluations, or about other persons' opinions. The information contributes wholly to fulfilling the consensus criteria and if accepted increases A's stability of attribution by that means. The question is whether A will accept B's information. On this point, A is confronted with a second type of problem of causal analysis, the one I mentioned earlier in which an observer makes an attribution of the action of another person. *Person B's message is itself an effect,* and A's problem is to attribute it, either to that part of their common environment under discussion (in which case it is considered valid), to B himself (his role,

desires, etc.), or to the situation or target (A himself, the particular circumstances).

In this attribution process, although they are used for a different purpose than in an earlier example (other person-environment attribution here *vs.* self-environment attribution in the earlier case), the familiar four criteria are again relevant:

1. *Distinctiveness.* Does B report this attribution only for this particular entity? If not, the cause of his report by this entity is cast in doubt.

2. *Consistency over time.* Does B report the same evaluation to all target audiences and not to A only? Does he make the same attribution in all situations, regardless of his motivation or circumstances? If not, there is indicated an attribution of his report to "irrelevant" causal factors such as A himself or B's own need state.

3. *Consistency over modality.* Does B report the same attribution when he thinks about the matter in different ways or when he approaches it with different observational methods? Is there any evidence in his presentation of his uncertainty, vacillation, indecisiveness, or weakness of response? These would be cues as to inconsistency and would suggest interference from other causal factors.

4. *Consensus.* Do other people agree with him? Does he convey information that others agree with him? (If so, this message itself creates a *further* problem of attribution and validation.) Or, does B provide information which implies that the reactions of other sensible persons might be different from his own? The latter would violate the consensus criterion for A, raise questions about whether B's reactions are really caused by the object, and reduce A's acceptance of B's assertions. (This may be illustrated by the ineffectiveness of a two-sided communication with less educated persons, in the study by Hovland, Lumsdaine, and Sheffield, 1949.)

The analysis of the validity of B's assertions can be summarized as in Figure 2. Person A (the "self" in this analysis) focuses on the "slice" of the data table produced by B (indicated by O_1) and tests for B's differential response to N, the consistency of his report (where the time categories refer to different situations or different target audiences), and his agreement with other observers of the same entities. I have left blank A's own "slice" of the data table (that corresponding to self) in order to emphasize his dependence upon B and other social sources of information. It is not uncommon, however, for A to have some reactions of his own to the entity, in which case the consistency between these and B's is an important determiner of the effect of B's message.

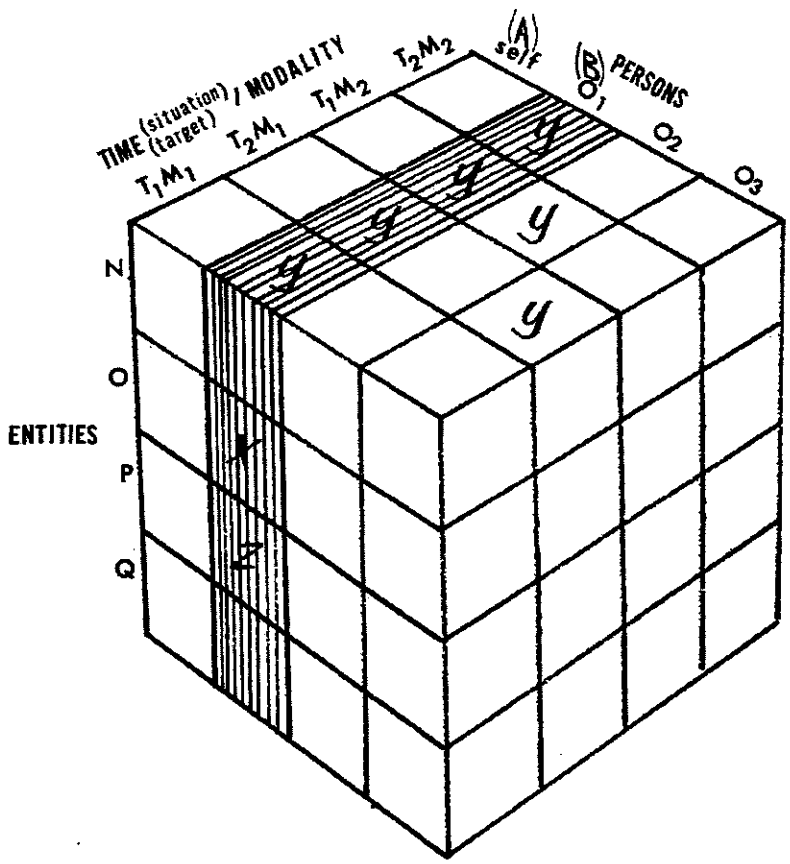

FIG. 2. Data pattern indicating attribution of B's message, Y, to entity N.

This reference to consistency between effects in A's own

"slice" of the data table and those in B's raises explicitly a problem that was finessed in the earlier example (e.g., in the reference to defining effects from "a given person's point of view"). What does it mean that another person's reaction to an entity is, as I see it and judge it, similar to or consistent with my own? This question is too complex to take up here, concerning as it does basic matters of interpersonal communication and comparison, ranging from emotional expression and perception to semantics and verbal labels. And the problem is not unique to the interpersonal comparisons. It also arises in the intrapersonal comparisons, in relation to consistency between effects experienced through different modalities (e.g., the estimate *vs.* the calculated answer to an arithmetic problem) and at different times (e.g., the vivid feelings of today *vs.* the recollection of last week's emotion). This is perhaps an appropriate place to mention a related point, that person B and other observers are simply special modalities that mediate A's contact with his world. (The reader may have noted in this regard that the tests applied to evaluate information obtained from B are the same in principle as the tests A would apply to evaluate any of his other modalities.) I consider them separately because of my special interest as a social psychologist in social influence and the social mediation of information.

The communicator factors usually considered relevant to A's acceptance of a persuasive communication are B's expertness and trustworthiness. These notions are readily reduced to attribution terms. *Expertness* can be defined as the communicator's contact with or mediation of the *relevant* external causal factors. An expert is one who, by virtue of his modes of interaction with the environment, is capable of attaining a high information level in his own attributions and (unless other causal factors interfere) of making attributional statements presumed to have high validity. *Trustworthiness* implies the *absence of irrelevant* causal factors (personal motives, role demands) in the person's statements. In view of the basic attributional relevance of these two concepts, it is perhaps no accident of method or approach that they have repeatedly appeared as defining the primary and exhaustive categories for the classification of communicator variables in research on persuasive communication.

In the case of this second method of influence, persuasion, the stability of A's subsequent attributions probably depends upon his keeping the message (evaluation, attribution) attached to its source and upon the stability of A's attribution of the message to the relevant external causes rather than to the communicator's irrelevant desires or situation. Some fascinating problems of memory are involved here, one of them having been revealed by Kelman and Hovland's (1953) work on remembering the content but not the source of a persuasive communication. The general point is that the various attributions and components involved in the initial acceptance of the message may not be retained over a period of time to the same degree. Consequently, there may be shifts over a period of time in the stability of attribution or even in the attribution itself.

The four criteria of attribution validity describe the different informational sources a person may have for a belief. The assembly of these criteria into an index of information level, modeled after an F ratio derived from an analysis of variance, implies how the several sources of information interact. There has long existed a dichotomy in social psychology between social reality and physical reality. The notions of temporal and modality consistency explicate the latter concept by specifying the nature of the information the person gains from physical reality, through his own direct interactions with the world. This specification has implications for the processes that are set into motion when a person encounters disagreement from other persons. The many experiments social psychologists have conducted on interpersonal influence show how a person faced with disagreement tests the consensus criterion by seeking opinions from still other persons. However, this research tells us very little about the consistency testing processes. For example, we know little about how the individual faced with disagreement tests his own temporal consistency, as by resampling his evidence, running over his arguments again, squinting and looking carefully at the stimulus (e.g., in the Asch experiment), or mentally checking and rechecking his solution to the problem. Nor do we know much about his checking across the various modalities available to him, as by changing his physical position and hence his perspective,

employing new observational aids, or seeking different ways of processing the problem data.

According to attribution theory, these processes will also entail various real or pseudo-experimental tests of the person who communicates inconsistent attributions. For example, in the event of sharp disagreement, A will wonder whether B is reacting to the same cause as he himself is. Person A will wonder whether, in Asch's (1952) terms, he is faced with a different judgment of the object or with a different object of the judgment. Person A may simply investigate whether this is the case, as by asking B to say more about the topic or to look again at the stimulus, so that A can judge for himself what it is that B is reacting to. Or, A may endeavor to control what it is that B reacts to, as for example, by a simple presence-absence test in which he checks B's reaction to some other entity and then has him return again to the disputed one for a new look at it. One is reminded here of the use of "filler" items and similar procedures used in psychological testing and lie-detection for catching the person off guard. Person A may also attempt to test B's motives by a presence-absence test. For example, if he suspects that B's assertion is caused by an interest in selling him something, A may try to remove this as a cause ("I'm not interested in buying; I can't afford it now") to see if the assertion varies accordingly, as it would if it were an effect of such an interest on B's part. Both of these last procedures are designed to ascertain the unique conditions under which B's message occurs and in that way to afford a basis for assessing its validity.

In the pursuit of satisfaction of the consensus criterion, persons in interaction with one another will often exercise influence in an attempt to produce unanimity of attribution. Attribution theory carries the important implication that mere compliance by a person initially expressing deviant opinions may not be satisfactory to his colleagues. If his deviant opinion has seriously reduced the stability of their causal attributions, his *mere compliance* (i.e., his response attributable to factors irrelevant to the causal question at hand) will not serve to restore the earlier level of stability. Only a change in opinion perceived as being a genuine reflection of the relevant cause (and now in line with

the colleagues' view) will satisfy their needs for stable attribution. Therefore, if a consensus is serving attributional purposes, genuine agreement will be sought. If it is merely in the service of group locomotion, on the other hand, genuine agreement may be far less important. Kenneth Ring's (1964) research is relevant to the former point. He finds that agreement with your ideas that is seen to be caused by your own high power or status does not provide you with information about the quality of your ideas. The agreement is seen as caused by your role and not by the ideas. Thibaut and Riecken's (1955) study of the perception of social causality has a similar implication.

It is one of the well documented hypotheses of social psychology that the pressures toward uniformity which are based in social reality are most effective when the physical reality checks are absent or meager. It has also been postulated, for example, by Festinger in his various theoretical statements (1950, 1954), that physical reality tests take precedence over social reality information, and certain experimental evidence can be interpreted as supporting this point of view. The implication is that the consistency criteria may be more important to the individual than the consensus criterion. Intuitively this seems reasonable, but the matter should be subjected to a careful attributional analysis. A partial explanation for predominance of the consistency criteria may be found in the point noted above, that influence mediated by the (perceived) consensus involves further attributional tests and is highly dependent upon attributions of the source. It might also be argued that a person generally has more information about his own motives as they enter into the attribution process and, therefore, is in a better position to exclude them as irrelevant causes of his reactions to entities than he is with respect to other persons' motives and their reactions. This raises interesting questions about the conditions under which priorities will be given to other persons' reactions rather than to one's own. These might be instances in which one's own causal role in the experienced effect is most to be suspected, as in evaluations of the self or of one's most highly cathected possessions.

Social Perception

The analysis of social influence has led us indirectly to the problem of perceiving the causes of another person's behavior (*viz.*, the behavior of a communicator). Jones and Davis (1965) have directly attacked this central problem in social perception. They employ attributional principles adopted from Heider to analyze the process of inferring another person's intentions from his actions. The authors define the task of their observer as that of identifying the intentions of the actor insofar as they depart from the normal or typical person in the same situation. In other words, the task is that of inferring the actor's *personal* attributes—the ways in which he is unique as a causal agent. The basic principles of their analysis are simple. It is first assumed that the observer believes the actor "was aware his action would have the observed effects" (p. 220) and has the "ability to bring about the effects observed" (p. 221). In other words, the effects may not be considered accidental. Under certain further conditions, the observer will then make a *correspondent inference* from the effects to the actor's intentions. By correspondent inference the authors mean that the effects and the attributes or intentions of the actor may be described in similar terms; e.g., behavior having the effect of dominating another person will be taken as a basis for inferring the actor's intention to dominate. The central hypothesis is that a correspondent inference will be made to the degree (a) there are *some but few* effects of the action that are unique (noncommon) to it as compared with other alternative actions available to the actor, and (b) these effects are *low* in assumed social desirability. The first part means that information is revealed about the actor's intentions only from the ways in which his chosen action differs from other actions he might have taken and that this information is less ambiguous the fewer such differences there are. The second part of the hypothesis means that the action yields information about the actor's idiosyncratic intentions only insofar as the effects are not those that people in general would have produced under similar circumstances. In brief, the fewer distinctive reasons a person might have for an action (assuming he has some) and the less these reasons

are widely shared in the culture, the more informative that action is concerning the characteristics of the person.

The matter of social desirability must be considered carefully. It is not implied that desirable effects are not relevant to the inference of intentions. Indeed, Jones and Davis note that we usually assume a person took an action out of regard for the good consequences it produced (assuming foresight and control, as we have) rather than with the intention of achieving its negative consequences. However, their purpose is to analyze the process of inferring *unique* intentions and "it is . . . clear that attribute-effect linkages based on universally desired effects are not informative concerning unique characteristics of the actor" (p. 227). In more general attributional terms, it might be said that effects widely produced by virtue of their desirability tell us more about the properties of external situations or states—their intrinsic desirability, demand characteristics, etc.—than about the properties of persons who act to produce them.

The reader will note some general parallels between the Jones and Davis analysis and my earlier one. (Indeed the reader may be puzzled that the correspondence is not more apparent, inasmuch as both analyses are alleged to stem from Heider. This is partly a reflection of the richness of Heider's ideas, but also an indication of the complexity of the problems at hand.) Let me try to explain the similarities and differences. To begin with a difference, the observer's focus in the two cases is essentially at opposite ends of the person-environment polarity. In my earlier analyses, dealing respectively with the self-environment and the other-environment problems, the person is concerned about the validity of an attribution regarding the environment. He applies the several criteria in an attempt to rule out person-based sources or "error" variance. In the problems specified by Jones and Davis, the observer has exactly the opposite orientation. He is seeking for person-caused variance (that caused by the particular actor under scrutiny) and in doing so, he must rule out environmental or situation-determined causes of variations in effects. He does this by looking for deviations from the general entity and situation trends, for example, as these are revealed in between-actor consistencies. This is the reason for the assumption that socially

undesirable effects are more informative than socially desirable ones.

In terms of the data table I have been using in this paper, the pattern of data fulfilling the Jones and Davis criteria is the one shown in Figure 3. The entries in the table represent various persons' actions as they are experienced by the self (i.e., in terms of their "effects"). As in the case of the persuasive communicator's message, these actions are themselves effects which must be subjected to causal analysis. The time dimension of the table includes

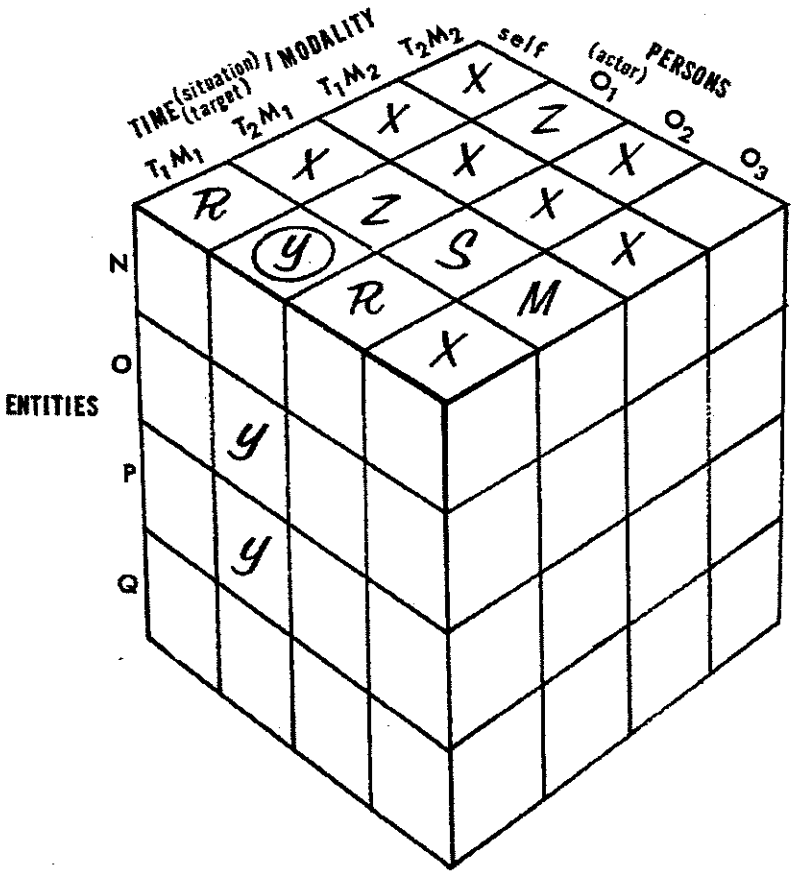

FIG. 3. Data pattern indicating attribution of effect Y to actor's intention.

(as in the analysis of the persuasive communication) the various situations in which the actions occur and different targets of these actions (the self and various others).

The Jones and Davis criteria lead to an inference of *intention* when the particular observed effect, Y, produced by O_1 at T_1M_1, deviates from the data trends for:

1. *Situations and targets.* The effect is not what most persons (or even this particular actor) produce in the same situation in relation to this particular target person, T_1. (Effects X and R are more common). This is essentially the criterion of low social desirability. The situation or person, or both, do not have "demand characteristics" such that most persons would produce the observed effect.

2. *The actor, O_1.* The effect is not to be accounted for by limitations in or peculiarities of the actor's available actions. Other actions that are known to be available to him have different effects (e.g., X and Z) even with the same target person and in the same situation (summarized by T_1). This deviation indicates that the observed effect is not to be attributed to the actor's "traits," habits, or response biases.

In brief, intention, as a reflection of a deliberate act of choice to produce the observed effect, is manifested in deviation of the observed effect from those attributable to situational demands and personal characteristics of the actor. From the present analysis, it might be added that the observed effect should also deviate from trends for:

3. *Entities.* The effect cannot be considered a reflection of the properties of some entity N that happens to be present.

4. *Modalities.* The effect is not a result of a particular way of the actor's interacting with, relating to, or viewing the self—a way that would induce any person to take an action with the same effect.

The second criterion above, deviations from the actor's "typical" effects, makes it clear why Jones and Davis assume that the observer believes the actor has foreknowledge and control. The actor's knowledge and ability enable him to choose among and employ his actions deliberately and with regard for their various effects. To the degree the effect of his selected action is different

from those of other actions he might have taken, the observer has evidence of the actor's willful intervention as a causal agent, exercising choice to cause a special effect.

This second criterion forms part of Jones and Davis' criterion of noncommon effects, specifying as it does that an inference as to intention requires that the observed effect deviate from those resulting from other actions of the actor. However, they also propose that *number* of noncommon effects is important—that the observer's inference about the actor's intention will be more firmly based the *fewer* are the effects that differentiate the selected action from the available alternative actions. The logic of this assumption is not clear. It seems reasonable to suppose that with fewer noncommon effects, it will usually be easier for an observer to label or rate the intention, but there seems no reason in principle why all or most of the various unique effects of an action may not be taken as evidence of the intention (or set of intentions) guiding the choice of that action. (In the present analysis, letters such as X, Y, and Z are used to indicate effects without regard to the number of identifiable components, and the term "effect" is used in a broad sense to include any single effect or set of consequences produced by an action, perceptual response, etc.)

As noted above, situational consistencies are introduced indirectly by Jones and Davis, by way of the concept of social desirability. To show the importance of social desirability, Jones, Davis, and Gergen (1961) present evidence from a study of observers' judgments of job applicants being interviewed for a job. Some applicants act in line with the requirements of the desired occupation and other applicants behave inconsistently with the requirements. The observers made more extreme attributions of the latter applicants and made the attributions with greater confidence. The interpretation is that inasmuch as most applicants would have acted in conformity with the situational requirements, it is only deviation from this conforming behavior that affords a basis for causal attributions to the individual person.

Jones and Davis' use of the criterion of social desirability and this experimental example contain an important point for our analysis: the observer may not need actually to examine variations in behavior over a number of persons. From his knowledge of social pressures, shared values, and situational demands, he may be able to make confident estimates about the amount of consensus of response to be expected. If made with sufficient confidence, whether correct or not, these estimates can supplant observation of the actual consensus and afford a basis for the attribution process. *This assumes,* of course, *that the individual has already made a firm attribution about the situation;* namely, that it has inherent properties such that most or all people respond to it in a given manner. The important general point is that once certain attributions are made, they become the basis for making further ones and they permit the individual to bypass some of the processes we have been describing here.

One aspect of my earlier analysis is missing from that of Jones and Davis, the criterion of temporal consistency. Their reason for omitting this from their scheme seems to be that they are exclusively interested in the inference of momentary intentions and not in "dispositional structures." This criterion would certainly be important in inferring enduring personal dispositions (e.g., the traits and response biases ruled out by criterion 2 above) and Jones and Davis imply as much. Temporal consistency might also be relevant in cases where the observer could not assume foresight and control. He might find it desirable to test this assumption by considering stability *vs.* change in actions after the actor has had opportunities over successive trials to observe and learn the effects of his actions.

Self Perception

Earlier, the question of attribution to the self *vs.* the environment was considered. I now wish to review several other approaches to the phenomena of self perception. Although these other approaches are consistent with this analysis of attribution theory, they pose new problems for the theory and suggest further directions for its development.

Recently Daryl Bem (1965, 1967) arguing from a radical behaviorist position, has proposed that the processes of arriving at inferences about the self are identical with those used in drawing inferences about others. For example, an observer judges

another's attitude on the basis of his behavior and the stimulus conditions under which it occurs. In the same way, a person judges his own attitude from observing his own behavior and taking account of the conditions under which it occurs. The latter assertion seems to contradict our everyday experience that we have direct and private access to our own attitudes. On the other hand, it is consistent with the present view that an attribution about the environment (an attitude would be a specific case) involves checking the consistency of one's responses, on successive occasions and over different modalities of interaction with the environment. In the sense that this checking process involves verbal responses, I would agree to the appropriateness of the comment from E. M. Forster, "How can I tell what I think till I see what I say."

A distinction should probably be made in regard to the above issue, a distinction between frequently or recently tested attributions and less well "rehearsed" ones. The former can probably be recovered directly in memory; i.e., I "know" in the sense of immediate recall what my attitude is toward the thing or what I believe its properties to be. Relevant to this point is Heider's warning against expecting the attribution process to be represented in experience. "Attributions may not be experienced as interpretations at all, but rather as intrinsic to the original stimuli" (p. 256).

The notion that the processes of self- and other-perception are similar is quite compatible with the present attribution analysis. Our examples have suggested that the same framework is used in making causal analyses whether the problem involves attribution to self *vs.* environment or to another person *vs.* his environment. This implies the same conclusion as Bem's, that given analogous information a person will draw the same conclusion whether in regard to himself or to another individual he is observing. However, this should not be taken to mean that the *same* information will always be available for the self as for another individual. To give a simple example, a person, being keenly aware of variations in the intensity of his own efforts, may observe correlated variations in his success on a series of tasks and attribute to himself a degree of control over this type

of task. On the other hand, an observer to whom the variations in "trying" may not be apparent may interpret the various performance outcomes as mere random variations in task difficulty and, consequently, may attribute no skill to the person. Another example would involve a difference between the actor and the observer (in the Jones and Davis problem of intention inference) in the set of alternative actions seen to be available to the actor. If different instances of TM (in Figure 3) are salient for the two persons, they will differ in the sets of actions (or effects) with which they compare the observed action, Y, and may differ, accordingly, in the intention to which they attribute it.

The specific problem Bem deals with in the context of self-perception *vs.* other-perception is that of inferring a person's "true" attitude from his verbal report or similar actions. This is a familiar problem at least insofar as it concerns inferring *another* person's true attitude. Earlier it was reported that the attribution analysis elicited by a communication of mere opinion by another person focused upon the assessment of the degree to which his opinion varied with variations in the entities but was consistent over time (situation, need state, etc.) and modalities, and with other persons' opinions. It is reasonable to expect the same analysis to be applied in the determination of one's *own* true attitude.

Bem employs B. F. Skinner's distinction between *tact* and *mand* in his conceptual analysis of this problem, and it appears that these categories of behavior have direct correspondence to our attributional analysis. Consider two classes of behavior (verbal response, or whatever), the first consisting of behavior that covaries with entities (and is consistently and consensually enacted) and the second consisting of behavior that varies with time (situation, need, target of the verbal behavior) rather than with entities. The first class corresponds to "tacts" which are responses under the discriminative control of some portion of the environment. Tacts consist of those portions of the behavioral repertoire that have been shaped by social training (largely language training) to vary only with entities; that is, to be uniquely distinctive for different entities, but highly consistent (in all the respects we have considered) for a given entity. They are "com-