

# The Role of Consistency in the Judgment of Stereotype-Relevant Behaviors

James L. Hilton  
William von Hippel  
*The University of Michigan*

*Two experiments examined the role that the consistency between a perceiver's stereotype and a target's behavior plays in the judgment of stereotype-relevant behaviors. In Experiment 1, consistency was varied through a manipulation of the objective properties of the behavior and its referent distribution. Ambiguous behaviors were assimilated to the stereotype when they were objectively consistent with the stereotype and were contrasted away from the stereotype when they were objectively inconsistent with the stereotype. In Experiment 2, perceived consistency was manipulated directly while objective consistency was held constant. Behaviors were identified either with members of random groups, with members of extended families, or with single individuals. Consistent with predictions, the tendency to assimilate ambiguous behaviors increased as expected consistency increased. These findings are discussed in light of recent work on contrast and assimilation and stereotyping.*

Numerous studies have documented the power that social stereotypes have to shape our judgments of those around us. Sometimes our stereotypes exert a distorting influence that leads us to assimilate our judgments to our stereotypes. Duncan (1976), for example, found that a shove was seen as playful if the protagonist was White and hostile if he was Black. At other times our stereotypes exert a distorting influence that leads us to contrast our judgments away from our stereotypes. Jussim, Coleman, and Lerch (1987), for example, found that well-dressed, well-spoken applicants were seen as more qualified if they were Black than if they were White. At still other times, our stereotypes appear to exert no distorting influence on our judgments of others. Nisbett, Zukier, and Lemley (1981), for example, found that when individuating information was available about a particular target, social stereotypes appeared to exert little influence over social judgments. But what determines when, and in what direction, stereotypes will affect people's judgments of others?

---

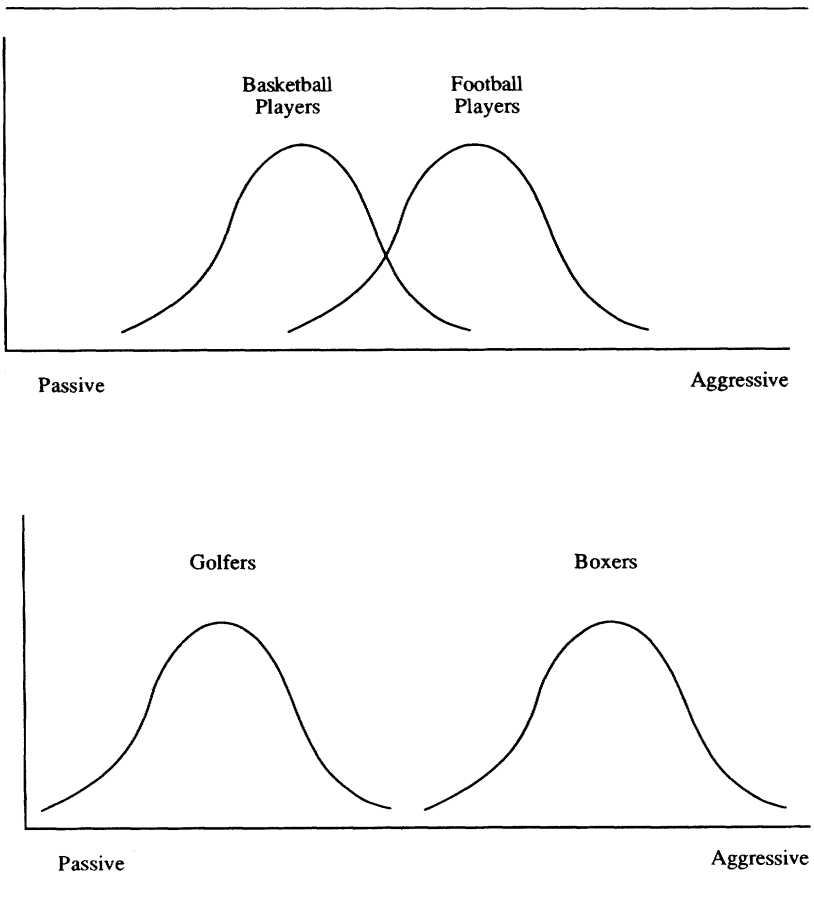
**AUTHORS' NOTE:** Preparation of this article was facilitated by a Rackham Faculty Research Grant and Fellowship from the University of Michigan and by a National Science Foundation grant (#BNS-8717784) to the first author. We gratefully acknowledge the assistance of Gene Burnstein, Nancy Cantor, Peter Ditto, Hazel Markus, Mel Manis, Claude Steele, and Bob Zajonc, who commented on earlier drafts of this article. Correspondence should be addressed to James L. Hilton, Research Center for Group Dynamics, Institute for Social Research, The University of Michigan, Ann Arbor, MI 48106.

*Personality and Social Psychology Bulletin*, Vol. 16 No. 3, September 1990, 430-448  
© 1990 by the Society for Personality and Social Psychology, Inc.

Most cognitive models of stereotyping (e.g., Allport, 1954; Bruner, 1957; Taylor, 1981) emphasize the role that categorization plays in the judgment of others. These models maintain that perceivers initially categorize individuals on the basis of some distinctive feature or information that is available about the person (e.g. "With that long hair, he's probably a liberal"; "With that kind of suit, she's probably a conservative"). Once the individual is categorized, perceivers then extrapolate from the information available from the category to make predictions about the person's behavior. For example, on the basis of what he or she knows about liberals, a perceiver might expect a man with long hair to vote in favor of a tax supporting parks and recreation and would be surprised if he or she were to learn that he had voted against the creation of an environmental cleanup fund. After making such predictions, perceivers then evaluate the consistency between the observed behaviors and the expected behaviors to determine whether the initial categorization was correct. Observed behaviors that are consistent with expected behaviors will confirm the initial categorization. Observed behaviors that are inconsistent with the expected behaviors can have several possible effects. In some instances, inconsistent information will lead perceivers to engage in an attributional analysis in an attempt to explain the inconsistency (Pyszczynski & Greenberg, 1981). In other instances, inconsistent information will lead perceivers to engage in stereotype revision (Weber & Crocker, 1983). And in still other instances, inconsistent information will lead perceivers to reject the initial categorization.

Whether any given behavior will be seen by the perceiver as consistent or inconsistent, however, is complicated by the fact that observed behaviors vary in the extent to which the information they provide is ambiguous. For our purposes, a behavior can be said to be ambiguous to the extent that the information that it provides is relevant to the categorization task but is compatible with more than one possible categorization. Crying, for example, will be ambiguous when one is attempting to determine whether another person is happy or sad, because crying is consistent with both great sorrow and joy (Trope, 1986).

In general, competing categorizations (e.g., liberal vs. conservative, man vs. woman) vary in the extent to which they generate overlap in the distributions of expected behaviors. Expected behaviors associated with some categorizations overlap considerably. Consider, for example, the range of behaviors associated with football players and basketball players. A perceiver who expects football players to behave more aggressively than basketball players may nevertheless recognize that some basketball players behave more aggressively than some football players. A hypothetical representation of the expected behaviors associated with football players and basketball players is contained in the top panel of Figure 1. Here the abscissa represents aggressiveness (more generally, any stereotype-relevant behavior) and the two curves represent the expected distributions of behaviors associated with football players and basketball players. Note that some of the behaviors are expected exclusively for football players and other behaviors are expected exclusively for basketball



**Figure 1** Hypothetical representation of the expected behaviors from overlapping and nonoverlapping categorizations.

players. These behaviors are relatively unambiguous because they are expected under only one categorization. Still other behaviors, however, fall between the means of the two distributions and are expected for both categorizations. These behaviors are relatively ambiguous because they are consistent with the expected behaviors associated with both categories.

The expected behaviors associated with other categorizations do not overlap. Consider, for example, the range of behaviors associated with boxers and golfers. Although boxers and golfers may share many nondiagnostic features (e.g., both may have short hair), they share few diagnostic features (e.g., golfers do not hit one another

in the ribs, and boxers do not congratulate one another after a good shot). Consequently, a perceiver who expects boxers to behave more aggressively than golfers may well expect every boxer to behave more aggressively than every golfer. A possible representation of the expected behaviors associated with boxers and golfers is contained in the bottom panel of Figure 1. Here it can be seen that the category labels are again associated with distributions of expected behaviors and that some of the behaviors are expected exclusively for boxers while other behaviors are expected exclusively for golfers. As in the previous example, these behaviors are relatively unambiguous because they are expected under only one categorization. Behaviors that fall between the means of the two distributions are again relatively ambiguous. In contrast to the previous example, however, behaviors that fall between the distributions are relatively ambiguous because they are *inconsistent* with the expected behaviors associated with both categories.

When consistency and ambiguity are considered jointly, two predictions can be made concerning the effects that competing categorizations will have on the judgment of an individual's behavior. First, if the observed behavior is unambiguous, the category label should have minimal influence on the judgment of the behavior (Herr, Sherman, & Fazio, 1983; Sherif & Hovland, 1961). Independent of whether the categories are overlapping or nonoverlapping, if a behavior falls exclusively within a single distribution of expected behaviors, there should be minimal distortion as a function of the category with which it is associated. A person who shouts another down, for example, is seen as assertive regardless of whether the person is male or female (Locksley, Borgida, Brekke, & Hepburn, 1980). Similarly, a target who is aware of a perceiver's cold expectancy is seen as warm regardless of whether the perceiver has a cold expectancy (Hilton & Darley, 1985). In instances like these, where the behavior is relatively unambiguous, the judgment of the behavior is essentially data driven (Norman & Bobrow, 1975), and the terms *confirmation* and *disconfirmation* refer to the consistency of the initial categorization, not to distortion of the observed event.

Second, if the observed behavior is ambiguous, the category label should have a distorting effect on the judgment of the individual's behavior. Moreover, the direction of the distortion should depend on the extent to which the behavior is consistent or inconsistent with the perceiver's expectations. With overlapping categorizations, ambiguous behaviors are consistent with the expected behaviors associated with both categorizations. The judgment of these behaviors should reflect distortion toward the activated stereotype. With nonoverlapping categorizations, ambiguous behaviors are inconsistent with the expected behaviors associated with both categorizations. The judgment of these behaviors should reflect distortion away from the activated stereotype.

Consistent with this analysis, Sherif and Hovland (1961) found that contrast and assimilation occur as a function of the distance between a behavior and its referent distribution. As the distance between an ambiguous event and its referent distribution increases, the tendency to contrast the event increases as well. Conversely, as the distance between an ambiguous event and its referent distribution decreases, the

tendency to assimilate the event increases. Similarly, Herr et al. (1983) and Herr (1986) found that priming with a similar category will produce assimilation whereas priming with a dissimilar category will produce contrast. Finally, Manis, Nelson and Shedler (1988) found that ambiguous behaviors are assimilated to a group stereotype when the behaviors are consistent with the behaviors of the other group members and are contrasted away from a group stereotype when the behaviors are inconsistent with the behaviors of the other group members.

But what is the best way to represent consistency? Previous research has examined consistency in terms of the objective features of the stimulus. Sherif and Hovland (1961), for example, manipulated consistency through the mechanism of distance; Herr et al. (1983) manipulated consistency through feature similarity. Although these lines of research make it clear that objective consistency (i.e., distance or feature similarity) plays a role in contrast and assimilation, as a psychological construct it may be useful to think about consistency in more subjective terms. Indeed, Sherif and Hovland (1961) were surprised to find that the relationship between objective consistency (i.e., distance) and contrast and assimilation sometimes breaks down. These breakdowns might be expected, however, if subjective consistency plays a mediating role in contrast and assimilation. That is, any factor that influences perceived consistency should interfere with the relationship between objective consistency and contrast and assimilation. Factors that increase the perceived consistency between an observed event and an expected event should lead to assimilation; factors that decrease the perceived consistency between an observed event and an expected event should lead to contrast.

In order to examine this issue, two experiments were conducted. The goal of the first experiment was to manipulate objective consistency and subsequently measure subjective consistency. If subjective consistency plays a role in contrast and assimilation, then not only should objective consistency translate linearly into subjective consistency, but subjective consistency should mediate the relationship between objective consistency and contrast and assimilation. The goal of the second experiment was to examine, in direct manner, the mediating role that subjective consistency plays in contrast and assimilation. In both experiments, we relied on a modified procedure developed by Manis and his colleagues (e.g., Manis, Paskewitz, & Cotler, 1986) that allows for the creation of highly specific stereotypes in the laboratory and enables the evaluation of subjects' judgments of subsequently encountered stereotype-relevant behaviors.

## EXPERIMENT 1

### Method

*Subjects.* Subjects were 71 students enrolled in introductory psychology. Their participation was in partial fulfillment of course requirements.

*Design.* During an induction phase, subjects rated 62 word definitions on the degree of pathology reflected in each definition. Thirty-one of the definitions were said to emanate from patients at Metropolitan Hospital and 31 were said to emanate

from patients at Central Hospital. In the nonoverlapping conditions, all 31 definitions associated with Metropolitan Hospital were normal and all 31 definitions associated with Central Hospital were pathological. In the overlapping conditions, 16 of the definitions associated with Metropolitan Hospital were normal and 15 were midrange, and 16 of the definitions associated with Central Hospital were pathological and 15 were midrange. Once subjects had completed this induction phase, they rated 18 additional definitions, which ranged across the continuum from normal to pathological. During this test phase, the definitions were labeled as having come from patients at Metropolitan Hospital for half of the subjects and Central Hospital for the other half of the subjects. These manipulations resulted in a 2 (nonoverlapping vs. overlapping induction series)  $\times$  2 (Metropolitan versus Central test series labels)  $\times$  3 (normal versus midrange versus high-pathology definitions during the test phase) factorial design, in which the first two factors were manipulated between subjects and the third factor was manipulated within subjects.

*Procedure.* Subjects were contacted by phone and scheduled to arrive in the lab in groups of up to six. Upon their arrival, subjects were seated individually in small booths containing a chair and a computer terminal. The computer presented the instructions and stimulus materials and allowed subjects to proceed at their own pace. The experimental session typically lasted 30 min.

At the outset, subjects were informed that the purpose of the experiment was to assess their ability to judge the pathology of definitions provided by patients from two imaginary hospitals (Central and Metropolitan). These two hospitals were described as general medical facilities, in which the patients could have been in any of a number of wards (e.g., cardiology, psychiatry, surgery). Subjects were told that their task would be to judge the degree of psychopathology reflected in the definitions the patients provided and to record their judgments on 9-point bipolar scales.

Once subjects had finished reading the instructions, the computer began the induction series. The induction series consisted of 62 definitions, half of which were said to emanate from Central Hospital and half from Metropolitan Hospital. The definitions were previously rated for the level of pathology they represented (Arnhoff, 1954) and ranged across the continuum from normal to pathological. Table 1 provides examples of normal, midrange, and pathological definitions.

In the nonoverlapping conditions, subjects rated 31 definitions from Metropolitan Hospital that were normal (range 1.0 to 4.3 on a scale of 1.0 to 11.0) and 31 definitions from Central Hospital that were pathological (range 7.7 to 11.0). This manipulation was designed to create in subjects the stereotype that patients from Metropolitan Hospital were uniformly normal whereas patients from Central Hospital were uniformly pathological. In the overlapping conditions, subjects rated 16 definitions that were normal and 15 definitions that were midrange (range 1.0 to 7.7) from Metropolitan Hospital and 16 definitions that were pathological and 15 that were midrange (range 4.3 to 11.0) from Central Hospital. This manipulation was designed to create the stereotype that patients from Metropolitan Hospital were normal to midrange in pathology whereas patients from Central Hospital were midrange to high in pathology.

**TABLE 1: Sample Definitions From the Induction Series**

<i>Degree of Pathology and Word</i>	<i>Definition</i>
High pathology	
Fable	Trade good sheep to hide in the beginning.
Mosaic	A stone place to put people, death.
Cushion	To sleep on a pillow of God's sheep.
Middle pathology	
Seclude	To hide, remove from probing eyes.
Spangle	Means of bright like star spangles.
Pewter	Something that don't smell so good.
Low pathology	
Cushion	A padded item used for comfort.
Gamble	Take a chance, a risk.
Bacon	Product of a pig.

Once the induction series was completed, subjects were told that their task had changed slightly. The instructions on the computer explained that all the subsequent definitions would originate from different patients at the two hospitals and that the subjects' task was now threefold. First, they were to report whether they thought each patient was pathological, recorded as a simple yes or no answer. Second, they were to indicate their confidence in this judgment, recorded on a 9-point bipolar scale. Third, they were to report whether they thought each patient in question was really interned at the hospital indicated, recorded as a simple yes or no answer.

Subjects used this procedure to rate definitions from patients who were identified as having come from either Metropolitan or Central Hospital. Half the subjects rated definitions from patients at Metropolitan and half rated definitions from patients at Central. These definitions consisted of 18 randomly presented sentences, identical in style to those presented in the induction series and spaced evenly across the continuum from normal to pathological. Once subjects had completed their ratings of the test series, they were debriefed, thanked, and dismissed.

## Results and Discussion

*Judgments of the targets.* Recall that subjects were expected to assimilate midrange items to the stereotype when the induction series exposed them to overlapping distributions and to contrast these same items away from the stereotype when the induction series exposed them to nonoverlapping distributions (Manis et al., 1988). Furthermore, subjects were not expected to show any distortion in their evaluation of either low- or high-pathology items. In order to test these predictions, we used a procedure adopted from Manis et al. (1986) in which the subjects' yes/no responses to the pathology question were combined with their associated confidence ratings to form a single scale. In creating this scale, a value of  $-1$  was assigned to "no" responses

**TABLE 2: Ratings of Targets' Pathology**

<i>Degree of Pathology and Induction</i>	<i>Hospital Label</i>	
	<i>Metropolitan (Normal)</i>	<i>Central (Path.)</i>
Low pathology		
Overlapping	4.3	4.4
Nonoverlapping	4.2	5.2
Midrange pathology		
Overlapping	8.0	9.3
Nonoverlapping	9.4	8.1
High pathology		
Overlapping	13.2	14.0
Nonoverlapping	13.9	14.4

and a value of +1 was assigned to "yes" responses. The confidence ratings were then multiplied by these numbers. All ratings were then converted to a positive scale by adding 9 to the product when positive and 10 to the product when negative. This transformation resulted in a bidirectional confidence scale in which 1 represents a high degree of confidence that the person is normal and 18 represents a high degree of confidence that the person is pathological. After creating this scale, we then computed a mean for the six ratings at each of the three levels of pathology and submitted the resulting index to a 2 (induction series)  $\times$  2 (test series label)  $\times$  3 (level of pathology) mixed-model analysis of variance (ANOVA).

As expected, a main effect for the pathology of the definitions emerged, indicating that low-pathology definitions were perceived as less pathological (4.5) than midrange definitions (8.7), which were, in turn, perceived as less pathological than high-pathology definitions (13.9),  $F(2, 66) = 270.8, p < .001$ . More important, the predicted three-way interaction emerged,  $F(2, 66) = 3.5, p < .05$ , with all 12 of the means in the predicted directions (see Table 2). Simple effects analyses revealed that, with low- and high-pathology definitions, neither the hospital label nor the induction series had any independent or interactive effects on subjects' judgments of the test items (all  $F_s < 1$ ). With midrange definitions, however, the predicted interaction between hospital label and induction condition emerged,  $F(1, 67) = 3.5, p < .07$ . As can be seen in Table 2, when the midrange items were consistent with the associated distribution, assimilation tended to occur. When midrange items were inconsistent with the associated distribution, contrast tended to occur. No other main effects or interactions emerged from the analysis (all  $F_s < 1$ ).<sup>1</sup>

*Recategorization.* Recall that, after evaluating the pathology of each individual in the test phase, subjects also indicated whether they believed that he or she was actually at the hospital specified. According to our hypothesis, subjects should have perceived low-pathology items as clearly consistent with low-pathology labels and clearly inconsistent with high-pathology labels. The opposite pattern of results should have

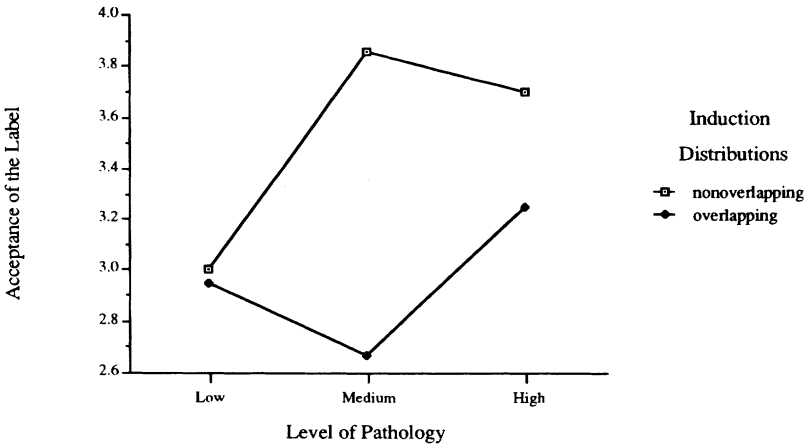


emerged with high-pathology items. Given that the level of pathology reflected in the definitions was manipulated orthogonal to the hospital labels associated with the definitions, half the labels for both the low- and high-pathology items should have been accepted, and half the labels should have been rejected. Moreover, the extent to which the induction distributions overlapped should have had little impact on the acceptance or rejection of the hospital labels for low- and high-pathology definitions. A different pattern of results should have emerged for midrange definitions. Midrange definitions were relatively consistent or inconsistent with the labels depending on whether the induction series consisted of overlapping or nonoverlapping distributions, respectively. Subjects should have perceived midrange items as consistent with either label under overlapping conditions and inconsistent with either label under nonoverlapping conditions. Consequently, the labels associated with midrange items during the test series should have been generally accepted under overlapping conditions and generally rejected under nonoverlapping conditions.

In order to examine these predictions, we assigned a value of 1 to subjects' acceptance of the label and a value of 0 to their rejection of the label. We then created an index by summing each subject's responses at each level of pathology. Because there were six definitions at each level of pathology, each of the three resulting values had a possible range from 0 to 6. We then submitted the resulting index to a 2 (induction series)  $\times$  3 (level of pathology) mixed-model ANOVA.

The ANOVA revealed a main effect for induction series, indicating that the labels were more readily accepted when the induction series consisted of overlapping distributions than when they consisted of nonoverlapping distributions,  $F(1, 69) = 6.9, p < .02$  (see Figure 2). The interpretation of this effect is qualified, however, by the presence of the predicted interaction between the induction series and the level of pathology,  $F(2, 68) = 4.6, p < .02$ . Simple effects analyses revealed that for high- and low-pathology items there was no evidence that the induction manipulation had any effect (both  $F_s < 1, n.s.$ ). For midrange items, however, subjects were more likely to accept the labels when the induction series consisted of overlapping distributions than when the induction series consisted of nonoverlapping distributions,  $F(1, 69) = 12.6, p < .001$ .

*An analysis of the role of perceived consistency.* If perceived consistency does indeed mediate contrast and assimilation, then the midrange items that subjects perceived as consistent with their associated stereotypes should have been assimilated to those stereotypes, and the midrange items that subjects perceived as inconsistent with their associated stereotypes should have been contrasted away from those stereotypes. Furthermore, this effect should have been independent of the particular stereotype with which the items were associated. In order to test these predictions, we examined subjects' tendencies to contrast or assimilate behaviors on the basis of their idiosyncratic acceptance or rejection of the hospital label accompanying each item. This was accomplished through the use of path analyses in which we examined both the influence of the induction series and the hospital label on acceptance of the label and the influence of the inductions series, the hospital label, and acceptance of the label on subjects' ratings of the pathology of the definitions. Conducting separate path

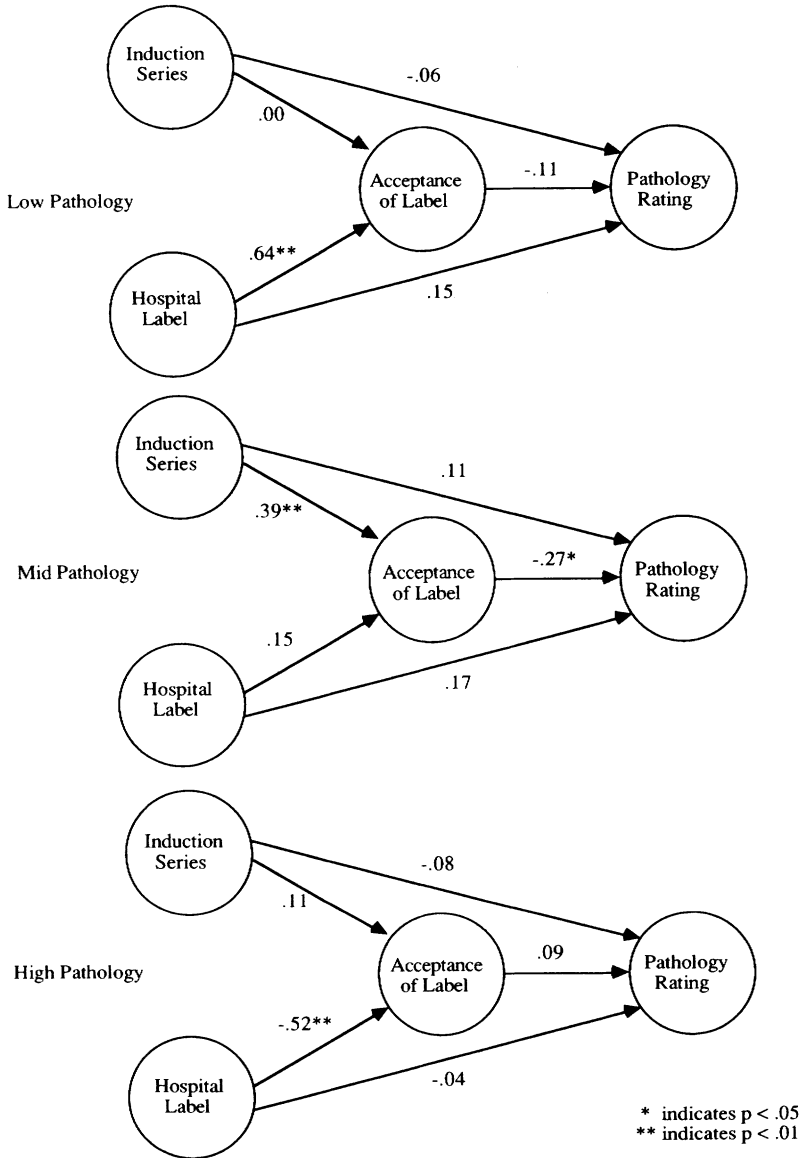


**Figure 2** Subjects' acceptance and rejection of the hospital labels as a function of category overlap and pathology of definitions. (Higher numbers indicate greater acceptance.)

analyses at each level of pathology removes the interaction effect that can be seen in Figure 2, so that it does not affect the analyses.

According to our predictions, the hospital label should have had a direct influence on acceptance of the label for both the low- and high-pathology items. Independent of the nature of the induction series, subjects should have accepted high-pathology labels for high-pathology definitions and rejected low-pathology labels for high-pathology definitions. The opposite effect should have occurred with low-pathology definitions. Furthermore, given the unambiguous nature of the low- and high-pathology items, neither hospital label nor type of induction series should have had any direct or indirect (via acceptance of the label) effect on subjects' ratings of the pathology of the definitions. With midrange-pathology items a different pattern should have emerged. For midrange-pathology definitions, the consistency between the induction series and the definitions should have had a direct influence on acceptance of the label. Independent of the hospital label provided, subjects should have accepted labels under overlapping distributions and rejected labels under nonoverlapping distributions. Furthermore, neither hospital label nor objective consistency should have had any direct influence on subjects' ratings of the pathology of the definitions. Rather, subjective consistency (acceptance of the label) should have mediated the effect of objective consistency on subjects' evaluations of the definitions.

As can be seen from Figure 3, all these predictions were confirmed. For high- and low-pathology definitions, the type of hospital had a direct effect on the acceptance of the label. Moreover, neither hospital label, type of induction, nor acceptance of the label had any direct or indirect effect on subjects' evaluation of the definitions.



**Figure 3** Path analyses of the effects of hospital label, induction distribution, and acceptance of the label on subjects' ratings of the targets' pathology.

For midrange-pathology items, however, objective consistency in the induction series influenced subjects' acceptance of the label, which, in turn, influenced their evaluations of the definitions.

The results from the path analyses suggest that perceived consistency may have mediated the contrast and assimilation effects found with ambiguous items. When subjects perceived moderately pathological definitions as consistent with their stereotypes, they tended to distort their evaluations in stereotype-confirming ways. When subjects perceived moderately pathological definitions as inconsistent with their stereotypes, they tended to distort their evaluations in stereotype-disconfirming ways. In the study that follows, we examine in a more direct manner the mediating role that perceived consistency plays in contrast and assimilation.

## EXPERIMENT 2

The three-way interaction from the first study supports the prediction that ambiguous behaviors are assimilated to consistent stereotypes and contrasted away from inconsistent stereotypes whereas unambiguous behaviors are neither assimilated nor contrasted in their evaluation. More important, the path analyses raise the possibility that factors other than objective consistency might mediate contrast and assimilation. If objective consistency has its effect through perceived consistency, then it may be possible to influence contrast and assimilation through a manipulation of perceived consistency while holding constant the objective components of the stimuli. If so, this would suggest that perceived consistency might be an important mediating variable in the contrast and assimilation process. In order to test this hypothesis, we chose to manipulate perceived consistency through the mechanism of group bondedness.

Srull (1981) has argued that groups vary in the extent to which psychological bonds tie group members together. Psychologically bonded groups are those from which perceivers expect a high degree of consistency among group members on some central dimension(s). Psychologically unbonded groups are those from which perceivers expect a low degree of consistency among group members on any central dimension. Consider two groups that differ in the extent to which they are bonded. One group consists of people who happen to be waiting at an airport for the same flight; another group consists of people who belong to the same fraternity. Among the members of the group waiting for the flight, we would not be surprised to find a wide range of opinions, personalities, and behaviors. We would not, for example, be terribly surprised to find that one member of the group is a Marxist and another is a staunch conservative. Among the fraternity brothers, however, we do expect consistency. We would be surprised to learn that a Marxist and a conservative belong to the same fraternity.

The difference in expected consistency between bonded and unbonded groups has two somewhat contradictory implications for the judgment of individual group members. First, in the absence of any other information, perceivers should expect greater variability in the behavior of members of unbonded groups. This characteristic of bondedness is captured in the aforementioned examples. Second, if the expected variability of members' behaviors is held constant, there should be greater psycho-

logical pressure to assimilate unexpected behavior to a bonded group than to an unbonded group. If, for example, we were to learn that a member of the group waiting for the flight had committed a murder, there would be no reason to attempt to reconcile his or her actions with the actions of the other group members. In fact, the onerousness of his or her deed might stand in rather stark contrast to the more normal behaviors of the other group members. If, however, we were to learn that a member of the fraternity had committed a murder, we might well try to reconcile his actions with the more normal actions of his fraternity brothers. We might, for example, see the murder as a crime of passion rather than as a cold-blooded killing.

The primary goal of the second study was to determine whether the tendency to see inconsistent behaviors in stereotype-consistent ways (i.e., to assimilate them to the activated stereotype) would increase as the expected consistency of behaviors among group members increased. To this end, we relied on the procedures outlined in Experiment 1, with three modifications. First, because we were interested in behaviors that were susceptible to distortion, only midrange definitions were used in the test series. Second, in order to examine the mediating role of consistency, the expected consistency between group members was varied. To create a high expectation of consistency, some subjects were told that the definitions emanated from two individuals.<sup>2</sup> To create a medium expectation of consistency, other subjects were told that the definitions emanated from two different families. To create a low expectation of consistency, still other subjects were told that the definitions emanated from two groups of people whose last names fell within the same half of the alphabet. Third, in order to hold the objective consistency of the behaviors constant, all subjects were exposed to the same, nonoverlapping induction distributions. Recall that these induction distributions led to contrast effects when the groups were composed of minimally bonded members of the same general medical facility. If increasing the expected consistency of a source increases the pressure toward assimilation, then contrast should emerge with a low-consistency source, and assimilation should emerge with a high-consistency source.

## Method

*Subjects.* Subjects were 87 students enrolled in introductory psychology. Their participation was in partial fulfillment of their course requirements.

*Procedure.* With few exceptions, the procedures followed in the current study were identical to those in the previous study. During an induction phase, subjects rated 60 word definitions on the degree of pathology reflected in each definition. In the low-consistency conditions, the sources of the definitions were described as groups of people whose last names begin with a letter in the same half of the alphabet (First Alphabet Group and Second Alphabet Group). In the medium-consistency conditions, the sources of the definitions were described as two different families (Barnes and Drake). In the high-consistency conditions, the sources of the definitions were described as two individuals (Bob and Mike). The induction phase always presented subjects with nonoverlapping distributions consisting of 30 pathological definitions associated with one source and 30 normal definitions associated with the other source.

During the test phase, subjects rated six more definitions that were identical in style to the definitions presented during the induction phase. The six definitions presented during the test phase were all located at the middle of the pathology continuum. Subjects were told that their task was to indicate whether or not they thought the definitions were pathological and how confident they were in this judgment. Half the subjects rated definitions from the pathological source, and half rated definitions from the normal source. Once subjects had completed their ratings of the test series, they were debriefed, thanked, and dismissed. These manipulations resulted in a 2 (normal vs. pathological labels in the test phase)  $\times$  3 (low vs. medium vs. high expectation of consistency) factorial design.

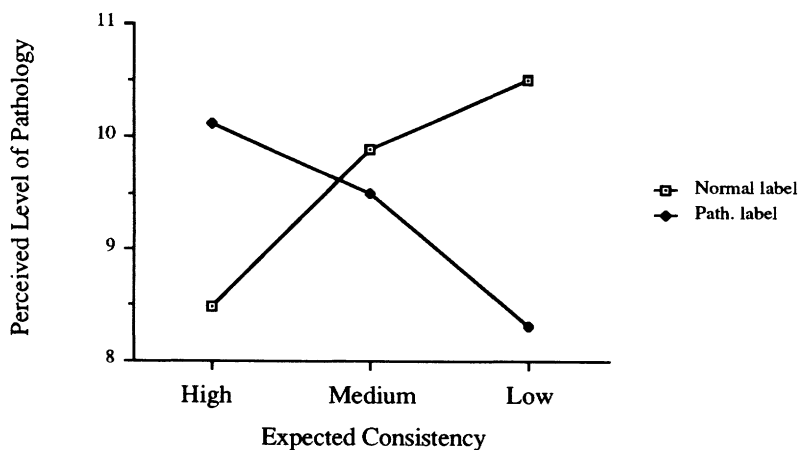
## Results and Discussion

In the current study, the primary hypothesis was that subjects would contrast behaviors away from the label when the items were associated with a low-consistency source and would assimilate behaviors toward the label when the items were associated with a high-consistency source. In order to test this hypothesis, we combined subjects' yes/no responses to the pathology question with the associated confidence ratings as in the previous experiment. We then computed a mean for the six ratings and submitted the resulting index to a 2 (test series label)  $\times$  3 (level of expected consistency) ANOVA.

As expected, the ANOVA revealed no main effects for either expectancy strength or test label (both  $F$ s  $< 1$ , n.s.). The predicted interaction between these two factors, however, was significant,  $F(2, 81) = 3.3, p < .05$  (see Figure 4).

Simple effects analyses revealed that, within the low-consistency conditions, definitions that were associated with the pathological source were seen as less pathological (8.3) than when they were associated with the normal source (10.5),  $F(1, 82) = 4.1, p < .05$ . Within the high-consistency conditions, the opposite pattern emerged, although the difference failed to achieve conventional levels of significance. Definitions associated with the pathological source were seen as more pathological (10.1) than definitions associated with the normal source (8.5),  $F(1, 82) = 2.5, p < .15$ . Within the medium-consistency source no label effects emerged,  $F(1, 82) < 1$ , n.s.<sup>3</sup>

Because all subjects rated the same induction and test phase definitions, a correlational analysis provides another way of examining these effects. If subjects' judgments of the test items were assimilated to their stereotypes, their ratings of the definitions presented during the test phase should be positively correlated with their ratings of the definitions presented during the induction phase. That is, if the subjects' judgments of the test items were assimilated to their stereotypes, their ratings of the definitions identified as having come from a pathological or a normal source during the test phase should be positively correlated with their ratings of the definitions identified as having come from the same source during the induction phase. However, if their judgments of the test items were contrasted away from their stereotypes, their ratings of the definitions identified as having come from a pathological or a normal source during the test phase should be negatively correlated with their ratings of the definitions identified as having come from the same source during the induction



**Figure 4** Subjects' ratings of pathology as a function of expected consistency of the source and type of label. (Higher numbers indicate greater pathology.)

phase. The results supported these predictions. In the high-consistency conditions, the correlation between the mean of subjects' ratings of the definitions presented during the test phase and the mean of their ratings of the definitions identified with the same source during the induction phase was  $r(29) = .46, p < .01$ . In the medium-consistency conditions, the correlation was  $r(28) = .12, n.s.$  In the low-consistency conditions, the correlation was  $r(27) = -.13, n.s.$  Fisher's  $z'$  transformation revealed that the correlation in the high-consistency condition was marginally different from the correlation in the medium-consistency condition,  $z = 1.37, p < .09$  (one-tailed), and significantly different from the correlation in the low-consistency condition,  $z = 2.26, p < .02$  (one-tailed).

Overall, the results from the second study demonstrate the potentially important role that perceived consistency plays in contrast and assimilation. Despite the fact that objective consistency was held constant, unexpected behaviors were assimilated when there was a high expectation of consistency from the source and were contrasted when there was a low expectation of consistency from the source. Thus, these results suggest that greater attention needs to be given to factors other than objective consistency that may play a role in contrast and assimilation. Although the objective consistency between the targets' behaviors and the behaviors of other group members can be offered as an explanation for the contrast and assimilation that resulted in the first study, the path analyses from that study and the pathology ratings from the current study suggest that perceived consistency, rather than distance, mediates contrast and assimilation. As noted earlier, Sherif and Hovland (1961) found that the relationship between distance and contrast and assimilation breaks down under certain conditions.

Specifically, they found that as materials become ego-relevant, distance no longer accurately predicts contrast and assimilation. But if perceived consistency is the mechanism through which contrast and assimilation work, then any factor that increases the perceived consistency between behaviors and stereotypes may increase the tendency toward assimilation. Consequently, as materials become ego-relevant, the relationship between distance and contrast and assimilation may not break down but may, rather, become biased toward assimilation. Future research examining the limits on contrast and assimilation and the relationship between expected consistency and contrast and assimilation could help clarify these issues.

Although the results from the current study support the notion that perceived consistency plays an important role in contrast and assimilation, it is important to note two potential limitations. First, the behaviors that were presented during the test phase were ambiguous in the sense that they were relevant to the categories that were formed during the induction phase but were equally compatible with both sources. Although the results from the first study suggest that perceived consistency will have minimal impact when the target's behavior is unambiguous, it is not clear what the effect of perceived consistency will be when ambiguity is operationalized in other ways. For example, it is not clear what effects perceived consistency will have on the judgment of behaviors that are irrelevant to the categories (Hilton & Fein, 1989; Locksley et al., 1980) or on the judgment of behaviors that are causally ambiguous (Jones & Davis, 1965; Kelley, 1967). Second, it is important to realize that, unlike subjects in the first study, subjects in the second study were not told that recategorization of the unexpected behaviors was possible. To the extent that recategorization is a viable alternative, the pressure to assimilate unexpected behaviors should diminish.

Finally, it is interesting to speculate on the relationship between our finding that high-consistency sources lead to assimilation and Srull's (1981) finding that these same sources lead to greater memory for inconsistent behaviors. It is possible that the memory advantage found by Srull reflects the cognitive work involved in the assimilation of these behaviors. Consistent with Srull's argument, the process of assimilating unexpected behaviors may lead perceivers to form more associative links to the unexpected behaviors. Future research that simultaneously measures contrast and assimilation and memory effects might help clarify the relationship between assimilation and memory.

## GENERAL DISCUSSION

At a general level, the current findings highlight the need to distinguish between labeling effects that reflect distortion on the part of the perceiver and those that do not. Sometimes behaviors are perceived as consistent or inconsistent with social stereotypes, and this evaluation reflects an accurate reading of the target's behavior. At other times behaviors are perceived as consistent or inconsistent with social stereotypes, and this evaluation is aided by perceptual distortion. Although the stereotyping literature has tended to concentrate on those instances in which the perceiver's expectations lead to distortion, we would like to suggest that equal attention needs to be given to those instances in which they do not lead to distortion.



Consider, for example, the implications that these results have for our understanding of the subtyping processes that emerge during stereotype revision. As described by Weber and Crocker (1983), subtyping involves the recategorization of expectancy-inconsistent information. As unexpected behaviors begin to accumulate, new subordinate categories are formed around the unexpected behaviors. They suggest, for example, that if we encounter a German professor who violates our stereotype of Germans as efficient people, we may create a subordinate category for German professors that includes inefficiency as a feature characteristic of the category. Through subtyping at the subordinate level, the original superordinate categorizations are maintained and the unexpected behaviors are simultaneously rendered less unexpected.

The current research contains some evidence for a process very similar to subtyping. Recall that when behaviors were inconsistent with the distribution of behaviors associated with the identified hospital but consistent with the distribution of behaviors associated with the other hospital, subjects indicated that the behaviors had been mislabeled, and the initial categorization had virtually no distorting effect. In other words, when subjects were able to recategorize inconsistent behaviors, the initial categorization no longer influenced their judgments of the target's behavior. These findings are relevant to the discussion of subtyping in two ways. First, they suggest that recategorization can sometimes occur at a superordinate level. A person who is initially categorized as a liberal but who subsequently endorses an economic policy that is strongly conservative, for example, may be recategorized as a conservative rather than subtyped as a particular kind of liberal. Of course, not all categorizations at the superordinate level are equally amenable to recategorization. Some superordinate categorizations are highly mutable. For example, it may be relatively easy to recategorize a liberal as a conservative or to recategorize a cold person as a warm person. Such categorizations depend more on social reality than on objectively determined physical reality. Other categorizations, however, will be relatively immutable. It should be difficult, for example, to recategorize a male as a female or to recategorize a Black as a White. As recategorization at the superordinate level becomes more difficult, subtyping should become more prevalent.

Second, the results suggest that when recategorization or subtyping occurs successfully (i.e., when a category at either the superordinate or the subordinate level can be found that uniquely captures the inconsistent behavior), the initial categorization will have little impact on the judgment of the individual. Once a subtype exists, its effects should be similar to the recategorization effects that emerged in the current research. A person who is successfully subtyped as a German professor should not be seen as more efficient as a function of having been initially categorized at the superordinate level as a German. If recategorization cannot occur, however, the initial categorization will exert a strong influence over the judgment of the behavior.

Before concluding, we should note two potential limitations of the current research. First, in both studies, subjects were explicitly led to form competing categories around two rather narrow distributions. It is not clear what effects will emerge when perceivers use their own, potentially less narrowly defined categories. Research on

the effects of category accessibility, however, suggests that the mere activation of a category will be sufficient to influence perceivers' subsequent judgments (Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1980). Second, in both studies, contrast effects emerged only when the test items fell outside the ranges of the induction distributions (i.e., when the test items were novel). This feature of the two experiments suggests that caution should be used in generalizing the contrast results. To the extent that perceivers are successful in creating new categories around unexpected behaviors, however, it may well be that contrast will tend to emerge only when the behaviors in question are novel.

Finally, the current results suggest that the power that stereotypes have to distort our judgments of others emerges from the joint action of ambiguity and perceived consistency. In both studies, ambiguous behaviors that were perceived as consistent with the activated stereotype were assimilated to the stereotype, and ambiguous behaviors that were perceived as inconsistent with the activated stereotype were contrasted away from the stereotype. Further attention to the roles of ambiguity and perceived consistency in the judgment of stereotype-relevant behaviors may better enable us to predict when stereotypes are most likely to have their pernicious effect.

#### NOTES

<sup>1</sup>The identical pattern of results emerged when the data were analyzed as simple proportions of yes and no responses, although the interaction effect was slightly weaker,  $F(2, 66) = 2.6, p < .08$ . Simple effects analyses revealed that, with low- and high-pathology definitions, neither the hospital label nor the induction series had any independent or interactive effects on subjects' perceptions of the test items: low pathology,  $F_s < 1$ ; high pathology, induction  $F(1, 67) = 1.4$  n.s., hospital label  $F(1, 67) = 2.1$  n.s., interaction  $F(1, 67) < 1$ , n.s. With midrange definitions, however, the predicted interaction between hospital label and induction condition emerged,  $F(1, 67) = 3.6, p < .07$ .

<sup>2</sup>In terms of expected consistency, the individual would seem to define one end of the continuum. We expect greater consistency among the behaviors of an individual than we do among the behaviors of different individuals from even the most strongly bonded group.

<sup>3</sup>The identical interaction emerged when the data were analyzed as simple proportions of yes and no responses,  $F(2, 81) = 4.1, p < .03$ . Simple effects analyses revealed that, within the low-consistency conditions, definitions that were identified as having come from the pathological source were seen as less pathological than when they were identified as having come from the normal source,  $F(1, 82) = 6.2, p < .02$ . Within the high-consistency conditions, the opposite pattern emerged, although the difference failed to achieve conventional levels of significance. Definitions identified as having come from the pathological source were seen as more pathological than definitions identified as having come from the normal source,  $F(1, 82) = 2.3, p < .15$ . Within the medium-consistency source no label effect emerged,  $F(1, 82) < 1$ , n.s.

#### REFERENCES

- Allport, G. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Arnhoff, F. N. (1954). Some factors affecting the unreliability of clinical judgments. *Journal of Clinical Psychology, 10*, 272-275.
- Bruner, J. (1957). On perceptual readiness. *Psychological Review, 64*, 123-152.

- Duncan, B. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*, 590-598.
- Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, *51*, 1106-1115.
- Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, *19*, 323-340.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141-154.
- Hilton, J. L., & Darley, J. M. (1985). Constructing other persons: A limit on the effect. *Journal of Experimental Social Psychology*, *21*, 1-18.
- Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. *Journal of Personality and Social Psychology*, *57*, 201-211.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). Orlando, FL: Academic Press.
- Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology*, *52*, 536-546.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192-238). Lincoln: University of Nebraska Press.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*, 821-831.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology*, *55*, 28-36.
- Manis, M., Paskewitz, J., & Cotler, S. (1986). Stereotypes and social judgments. *Journal of Personality and Social Psychology*, *50*, 461-473.
- Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*, 248-277.
- Norman, D. A., & Bobrow, D. G. (1975). On data limited and resource limited processes. *Cognitive Psychology*, *7*, 44-64.
- Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, *40*, 31-38.
- Sherif, M., & Hovland, C. T. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven, CT: Yale University Press.
- Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 440-463.
- Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, *38*, 841-856.
- Taylor, S. (1981). A categorization approach to stereotyping. In D. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, NJ: Lawrence Erlbaum.
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239-257.
- Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, *45*, 961-977.