Title
Attentional Processes in Stereotype Formation: A Common Model for Category Accentuation and Illusory Correlation

Journal
Journal of Personality and Social Psychology, 96(2)

ISSN
0022-3514

Authors
Sherman, Jeffrey W
Kruschke, John K
Sherman, Steven J
et al.

Publication Date
2009-02-01

DOI
10.1037/a0013778

Peer reviewed

Journal of Personality and Social Psychology
2009, Vol. 96, No. 2, 305–323

0022-3514/09/$12.00 DOI: 10.1037/a0013778

Attentional Processes in Stereotype Formation: A Common Model for Category Accentuation and Illusory Correlation

Jeffrey W. Sherman
University of California, Davis

John K. Kruschke, Steven J. Sherman, and Elise J. Percy
Indiana University Bloomington

John V. Petrocelli
Wake Forest University

Frederica R. Conrey
Synovate

Stereotype formation may be based on the exaggeration of real group differences (category accentuation) or the misperception of group differences that do not exist (illusory correlation). This research sought to account for both phenomena with J. K. Kruschke's (1996, 2001, 2003) attention theory of category learning. According to the model, the features of majority groups are learned earlier than the features of minority groups. In turn, the features that become associated with a minority are those that most distinguish it from the majority. This second process is driven by an attention-shifting mechanism that directs attention toward group–attribute pairings that facilitate differentiation of the two groups and may lead to the formation of stronger minority stereotypes. Five experiments supported this model as a common account for category accentuation and distinctiveness-based illusory correlation. Implications for the natures of stereotype formation, illusory correlation, and impression formation are discussed.

Keywords: stereotype formation, social categorization, categorization and attention, category accentuation, illusory correlation

Tajfel's experiments (Tajfel & Wilkes, 1963) on category accentuation and Hamilton's demonstration of the *distinctiveness-based illusory correlation* (Hamilton & Gifford, 1976) are the two seminal findings in the development of the social-cognitive approach to understanding stereotype formation. Whereas category accentuation effects highlight the exaggeration of real intergroup differences as the basis for stereotype formation, the illusory correlation shows that stereotypes may be formed in the absence of real group differences. Research on the two effects has largely proceeded independently, and they have been explained by different mechanisms. The aim of the present work was to place category accentuation and illusory correlation into a common theoretical framework, showing that both can be explained by Kruschke's (1996, 2001, 2003) attention theory (AT) of category learning. In so doing, this research also reconciles what have been viewed as incompatible accounts of the illusory correlation. Beyond integrating past research on stereotype formation, AT also suggests a number of novel predictions that were tested in this research.[1]

Jeffrey W. Sherman, Department of Psychology, University of California, Davis; John K. Kruschke, Steven J. Sherman, and Elise J. Percy, Department of Psychological and Brain Sciences, Indiana University Bloomington; John V. Petrocelli, Department of Psychology, Wake Forest University; Frederica R. Conrey, Synovate, Washington, DC.

Thanks to Cindy Pickett for her helpful comments on this research. This research was supported by National Institute of Mental Health Grant MH59774 to Jeffrey W. Sherman and by National Science Foundation Grant BCS-9910720 to John K. Kruschke.

Correspondence concerning this article should be addressed to Jeffrey W. Sherman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616. E-mail: jsherman@ucdavis.edu

Category Accentuation: Stereotype Formation as the Exaggeration of Real Group Differences

Tajfel's pioneering research showed that the division of graded stimuli into discrete categories exaggerates perceptions of the stimuli in the two categories. In the classic example, the placement of a category boundary between lines of varying length caused the lines in the long category to be judged as longer and the lines in the short category to be judged as shorter than when no category boundary was provided (Tajfel & Wilkes, 1963). Categorization exaggerates not only perceived differences between categories but also perceived similarities within categories. The accentuation of both between-category differences (e.g., Corneille & Judd, 1999; Eiser, 1971; Krueger & Rothbart, 1990; Queller, Schell, & Mason, 2006) and within-category similarities (e.g., Krueger & Clement, 1994; McGarty & Penny, 1988; McGarty & Turner, 1992) has been demonstrated in the perception of social groups. These effects contribute to the development of group stereotypes. Indeed, research has shown that stereotypes are most likely to be formed

[1] The present research focused specifically on the formation of novel stereotypes, not on the effects of existing stereotypes (e.g., *expectancy-based illusory correlations*; Hamilton & Rose, 1980). The effects of existing stereotypes on attention, encoding, judgment, and memory processes during impression formation have been well documented (for a review, see Hamilton & Sherman, 1994).

around attributes for which intergroup differences are large and intragroup differences are small (e.g., Ford & Stangor, 1992).

A variety of mechanisms have been shown to contribute to these accentuation effects. First, as detailed in Tajfel's original research (Tajfel & Wilkes, 1963), perceptions of individual category exemplars (or group members) may be biased by category boundaries. However, even when the objective nature of the exemplars prevents variation in their interpretation (e.g., the category members have discrete values, such as numbers or the presence vs. absence of a key feature), perceptions of the categories as a whole may still be accentuated. For example, category members that heighten between-category differences and within-category similarities may be attended to more carefully, given greater weight in judgments, or remembered more easily (Krueger & Rothbart, 1990; Krueger, Rothbart, & Sriram, 1989). Thus, perceptions of individual category members need not be exaggerated for category accentuation to occur. An important feature of each of these mechanisms is that they assume real differences between the categories in question and that these processes serve to accentuate those differences.

Illusory Correlation: Stereotype Formation in the Absence of Real Group Differences

The distinctiveness-based illusory correlation describes a phenomenon whereby observers perceive an association between distinctive groups and distinctive behaviors when, in fact, no such relationship exists (Hamilton & Gifford, 1976; for a review, see Stroessner & Plaks, 2001). In the modal demonstration of the effect, two groups (Group A and Group B) are described by a series of positive and negative behaviors. There are twice as many members of Group A as Group B, and there are more positive than negative behaviors. For example, in the original experiment, members of Group A performed 18 positive and 8 negative behaviors, whereas members of Group B performed 9 positive and 4 negative behaviors. Thus, Group B and negative behaviors both are distinct because of their infrequency.

Although Group A performs more behaviors than Group B, the ratio of positive to negative behaviors is the same in both groups. Consequently, there is no relationship between group membership and desirability. Nevertheless, participants perceive Group A to be more favorable than Group B. This effect is reflected in a number of findings. First, participants rate Group A more favorably than Group B on trait-rating tasks. Second, participants often overestimate the numbers of negative versus positive behaviors performed by Group B. Finally, participants misattribute Group A's negative behaviors to Group B (for a review, see Stroessner & Plaks, 2001).[2]

These effects have drawn intense interest because they demonstrate the formation of differential group impressions in the absence of real group differences. In addition, there is a direct parallel between the illusory correlation and the perceptions of majority and minority social groups outside of the lab. As in the illusory correlation, members of minority groups are numerically distinct, as are negative behaviors (e.g., criminal acts). Thus, the illusory correlation may serve as a model of how minority groups come to be viewed more negatively than majority groups in societies even if their behavior does not warrant such evaluations.

A variety of explanations have been offered for the illusory correlation. Each of the explanations has received empirical support, suggesting that the effect is multiply determined. Indeed, nothing in the current article is meant to suggest that there is one and only one factor in the development of illusory correlations. What follows are brief descriptions of the most prominent accounts.

The original account offered by Hamilton and Gifford (1976) is based on the notion that distinctive information draws attention. Because Group B members and negative behaviors are both numerically distinct, the combination of Group B members performing negative behaviors will be particularly salient (compared with any other group–behavior combination). The greater attention given to the distinctive negative Group B behaviors creates the misperception that Group B is, in fact, less favorable than Group A.

In contrast, Rothbart (1981) argued that because the Group A positive behaviors are the most common ones, they would be the most accessible in memory, resulting in more favorable judgments of Group A than Group B. In this case, it is the positivity of Group A rather than the negativity of Group B that drives the effect.

Smith (1991) explained the illusory correlation using Hintzman's (1986) memory-trace model. According to this account, the illusory correlation is based on the ways that group behaviors are retrieved and aggregated during the judgment process. The key feature of the model is that retrieval and aggregation of group behaviors are sensitive to differences rather than ratios in the numbers of positive and negative behaviors describing each group. According to this argument, Group A is evaluated more favorably than Group B because the difference in the numbers of positive and negative behaviors is greater for Group A ($18 - 8 = 10$) than for Group B ($9 - 4 = 5$).

Fiedler (1991; see also van Rooy, van Overwalle, Vanhoomissen, Labiouse, & French, 2003) argued that the illusory correlation results from regression to the mean following information loss. Participants do not perfectly encode the ratios of positive to negative behaviors in Groups A and B. As a result, when judging the favorability of the groups, estimates of the prevalence of positive and negative behaviors regress to the mean of all behaviors within each group (i.e., the average of the numbers of positive and negative behaviors). This leads to an overestimation of negative behaviors and an underestimation of positive behaviors. However, because Group B is smaller than Group A, the true ratio of positive to negative behaviors is learned less well in Group B than Group A. Consequently, there is more extensive regression to the mean in judgments of Group B than Group A, leading to a greater underestimation of the ratio of positive to negative behaviors for Group B. In sum, illusory correlation results from greater extraction of information about Group A than Group B, leading to greater regression to the mean in perceptions of the positivity/negativity of Group B.

Finally, McGarty and his colleagues (e.g., McGarty, Haslam, Turner, & Oakes, 1993; for a review, see McGarty & De la Haye, 1997) argued that the illusory correlation is not, in fact, illusory. They reasoned that the greater difference in the numbers of posi-

[2] Note that these effects are due to the numerical distinctiveness of the negative behaviors rather than to their negativity per se. Thus, the effects are reversed when the majority of the behaviors are negative rather than positive. In this case, Group B is perceived to be more positive than Group A (Hamilton & Gifford, 1976).

tive versus negative behaviors for Group A (18 − 8 = 10) than for Group B (9 − 4 = 5) constitutes a real group difference—that Group A is objectively more favorable than Group B. They further suggested that participants are motivated to differentiate the two groups and therefore seek to amplify this real group difference via category accentuation processes, such as biased perceptions of group behaviors (e.g., Berndsen, Van der Pligt, Spears, & McGarty, 1996) and other confirmatory hypothesis-testing strategies (e.g., Berndsen, McGarty, Van der Pligt, & Spears, 2001). Thus, in this view, the illusory correlation is a special case of the broader phenomenon of category accentuation explored by Tajfel and others.

Integrating Category Accentuation and Illusory Correlation: Attention Theory

As summarized above, wholly different processes have been posited to account for stereotype formation that is based on the exaggeration of real group differences and stereotype formation that is based on perceptions of illusory group differences. McGarty and his colleagues (McGarty & De la Haye, 1997; McGarty et al., 1993) posited the same mechanisms for category accentuation and illusory correlation but did so because they viewed the illusory correlation as reflecting the accentuation of real group differences. The purpose of the current research was to test Kruschke's (1996, 2001, 2003) AT model of category learning as a common explanation of stereotype formation based on both real and illusory group differences. The essential message of the AT analysis is that a common set of processes can produce both category accentuation and illusory correlation effects regardless of whether or not there are real differences between the groups in question.

The goal in this research was not to show or argue that AT can provide an exhaustive account of every finding from the category accentuation and illusory correlation literatures. It bears repeating that both effects appear to be multiply determined, and no single explanation (including AT) has been able to account for every data point. Rather, the purpose here was to show via AT that the exact same processes are sufficient to produce category accentuation and illusory correlation regardless of the status of group differences as real or not. In addition, AT makes a number of predictions about stereotype formation that are not made by existing models and cannot be accounted for by those models. Thus, AT not only offers a new, integrative account of known effects but offers a number of novel predictions as well.

Attention Theory and the Inverse Base-Rate Effect

AT was developed, in part, to account for the *inverse base-rate effect* in human learning. In the original demonstration of the effect (Medin & Edelson, 1988), participants were asked to diagnose different diseases from patterns of symptoms. On each trial of the learning sequence, a list of symptoms was presented, and participants were asked to diagnose the hypothetical patient as having one of several possible fictitious diseases. After each trial, participants were told the correct diagnosis. The basic design involved a pair of diseases, designated C (for common) and R (for rare), which occurred with a 3:1 ratio (usually, multiple pairs are presented to increase the difficulty of learning). During training, every instance of Disease C had two symptoms, labeled I and PC, and every instance of Disease R had two symptoms, labeled I and PR. PC and PR were perfect predictors of Diseases C and R—PC always predicted C and never R; PR always predicted R and never C. I was an imperfect predictor of the two diseases in that all cases of both C and R were associated with that symptom. For example, Disease C might always have the symptoms of a headache and a rash, whereas Disease R might always have the symptoms of a stomachache and a rash.

Following training, participants were tested with combinations of symptoms not shown during training. When tested with the ambiguous symptom I alone, people tended to choose the common disease, consistent with the base rates (during training, I appeared with C 3 times as often as it appeared with R). When tested with the combination I + PC + PR, people again tended to choose the common disease, though to a lesser extent. However, when presented with the conflicting symptoms PC + PR, participants tended to choose the rare disease, contrary (or inverse) to base rates. This inverse base-rate effect has been demonstrated in many other contexts and is not restricted to particular base rates or procedures (e.g., Kruschke, 2001).

AT explains the effect as follows: During training, people learn about frequent categories before they learn about rare categories. As such, participants first learn that Symptoms I and PC are typical of Disease C because that case occurs with high frequency (see Figure 1). Subsequently, they learn the rare disease, R. They discover that the shared symptom I is a misleading predictor of Disease R because it already is associated with Disease C. As a result, attention shifts away from I and is focused on the distinct symptom of R, PR. The purpose of this is to preserve and protect previous learning and to accelerate new learning. In fact, this attention-shifting mechanism causes PR to be more strongly associated with Disease R than PC is associated with Disease C. In effect, when learning about Disease R, attention is focused primarily on a single, distinctive symptom (PR), whereas when learning about Disease C, attention is divided between the PC and I symptoms.

When tested with I, people choose Disease C because I has been first associated with that disease. However, when tested with PC + PR, people choose Disease R because the symptom list contains the one key distinctive symptom of Disease R but only one of the two typical symptoms of Disease C. Moreover, because of the additional attention given to the distinct symptom PR, it is more strongly associated with Disease R than PC is associated with Disease C. Finally, when tested with symptoms I + PC + PR, people show a preference for Disease C because two of the three symptoms are typical of that disease. However, this preference

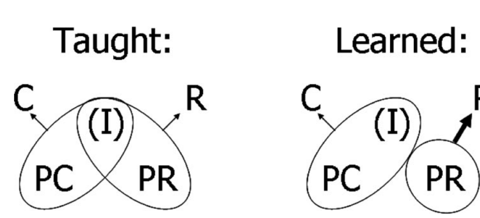

Figure 1. Left: The core design of the inverse base-rate effect. C and R denote common and rare diseases, respectively. PC is a perfect predictor of the common disease; PR is a perfect predictor of the rare disease. I is an imperfect predictor. Right: Depiction of what is learned according to attention theory.

may be muted because the numerical advantage of symptoms for Disease C is offset by the strength of the relationship between PR and Disease R.

Both behavioral research and formal connectionist models have shown AT to effectively explain the inverse base-rate effect (Kruschke, 1996, 2001), as well as other related learning phenomena, such as *base-rate neglect* (Gluck & Bower, 1988; Kruschke, 1996), *associative blocking,* and *highlighting* (Kamin, 1968, 1969; Kruschke, Kappenman, & Hetrick, 2005). Specifically, as category representations develop, the increased attention directed at exemplars that confirm category differences accentuates perceived differences between categories and similarities within categories. It is this attentional accentuation that causes a category member possessing both PC and PR attributes to be assigned to the rare category, despite base rates to the contrary (i.e., the inverse base-rate effect). There also is evidence from eye-tracking data that as category knowledge develops, attention shifts toward features of exemplars that accentuate between-category differences and within-category similarities (Kruschke et al., 2005) and that these shifts in attention predict categorization judgments.

Attention Theory and Stereotype Formation

It should be apparent that the learning and attention-shifting mechanisms in AT that allow people to distinguish categories from one another bear a strong resemblance to the category accentuation processes studied in the wake of Tajfel's research. For example, people attend most carefully to social category members that heighten between-category differences and within-category similarities, give those members greater weight in judgments, and remember those instances most easily (e.g., Krueger & Rothbart, 1990; Krueger et al., 1989). Perceivers also demonstrate biased perception of individual category members, such that greater attention is paid to features that assimilate them to their own group and contrast them away from other social categories (e.g., Tajfel & Wilkes, 1963). AT-like attentional accentuation processes are certainly consistent with and, in some cases, may directly underlie these findings.

However, one important difference between AT and models of social category accentuation is that AT does not require that there be real differences between the categories in question. Two categories may be described identically but still produce differentiated representations as long as one category is learned before the other. Indeed, any factor that causes one category to be learned prior to another (e.g., frequency of exposure, group size) will lead to attentional accentuation and to different and accentuated impressions of the categories (e.g., group stereotypes). The first category will be associated with its most common attributes, and driven by attentional accentuation, impressions of the second category will form around those features that most clearly differentiate it from the first category. Thus, AT provides an account not only of how categories are differentiated from one another but of which particular features come to characterize those categories.

These same processes also may produce the distinctiveness-based illusory correlation. According to AT, because Group A is larger than Group B, people learn about Group A first. Because positive behaviors are more frequent than negative behaviors, the impression formed of Group A is a positive one. Subsequently, in forming impressions of Group B, it must be the negative behaviors (the only remaining behaviors) that distinguish it from Group A and receive particularly close attention. Thus, to distinguish Group B from Group A, perceivers focus attention on Group B's negative behaviors. In this case, it is the distinctiveness of the negative behaviors vis-à-vis the existing impression of Group A that draws attention rather than their raw numerical distinctiveness, as described by Hamilton (Hamilton & Gifford, 1976). In a sense, AT's account of illusory correlation is a combination of the category accentuation approach favored by McGarty and the distinctiveness approach forwarded by Hamilton. According to AT, people do attempt to differentiate the two groups but do so, in part, by focusing on the Group B–negative behaviors because of their contextual distinctiveness. By this explanation, whether or not the differences between Group A and Group B are real is irrelevant. All that matters is that the first impression formed is a positive impression of Group A. AT also offers a number of novel predictions about the illusory correlation (to be detailed below) that cannot be explained by any existing model.

Overview of Experiments

We conducted five experiments to examine the possible application of AT to stereotype formation processes. In Experiment 1, we attempted to replicate the inverse base-rate effect with social categories described by trait attributes. In Experiment 2, we performed a variation on the standard distinctiveness-based illusory correlation paradigm to provide distinct support for the AT explanation of the effect. In Experiment 3, we showed that the inverse base-rate effect and distinctiveness-based illusory correlation could be produced simultaneously from a single procedure, lending support to the idea that the same mechanism underlies both phenomena. In Experiment 4, we further examined AT in the context of category accentuation. Finally, in Experiment 5, we directly tested the attention-shifting component of AT within an illusory correlation paradigm. Together, the five experiments provide strong support for the viability of AT as a general model of stereotype formation.

Experiment 1

Overview and Predictions

Experiment 1 was a conceptual replication of the standard inverse base-rate effect design described above, replacing symptoms and diseases with trait attributes and social groups. The purpose of this study was to demonstrate the effects of attentional accentuation on stereotype formation in a case in which there were clear, real group differences. The inverse base-rate effect is a category accentuation effect in that the processing of information about two categories is skewed in a way that serves to decrease the perceived similarity of the categories. This is demonstrated by the fact that exemplars possessing both the PC and PR attributes are assigned to the rare category, despite base rates to the contrary. Were there no accentuation process, then, presumably, those exemplars would be assigned to the common category, according to the base rates. There is now considerable evidence that this effect is produced by the attentional accentuation process described by AT (Kruschke, 1996, 2001; Kruschke et al., 2005). As such, the demonstration of an inverse base-rate effect in the present study would provide preliminary evidence that AT can account for the accentuation of real differences between social categories (i.e., stereotype formation).

Common (majority) and rare (minority) group members were described by perfect and imperfect predictor traits. For each group, there was a unique perfect predictor (i.e., a trait that always characterized that group and only that group: PC for the common group and PR for the rare group) and an imperfect predictor (i.e., a trait I that characterized every member of both groups). After learning about the group members, participants were presented with new group members who possessed novel combinations of traits and were asked to categorize these targets into one or the other group. We had four primary predictions. First, during initial training, we expected that participants would learn the features of the majority group before they learned the features of the minority group because the majority group was encountered more frequently. Second, during testing, we expected that, when novel group members possessed only the imperfect predictor (I), participants would be more likely to assign that person to the majority than the minority group. Third, when novel group members were described as possessing both of the perfect predictors (PC and PR), we expected that participants would be more likely to assign them to the minority than the majority group, replicating the inverse base-rate effect. Finally, we expected that, when novel group members possessed all three traits (PC, PR, and I), the inverse base-rate effect would be muted or even reversed because of the influence of the imperfect predictor's association with the common group.

Method

Participants. For their participation, 38 students at Northwestern University (Evanston, IL) were given partial course credit in an introductory psychology course. Participants were run in sessions of 1-4 people.

Materials and procedure. Participants were asked to engage in a categorization task. On each trial, they were presented with a name and two descriptive traits (e.g., John; reliable and stingy) and were asked to assign the name to one of four groups (Is this a member of Group F, G, H, or J?). Groups F and H shared an imperfect predictor, as did Groups G and J. Thus, each participant learned two replications of the inverse base-rate design with two pairs of groups. Members of Groups F and H were characterized by the traits reliable, friendly, and stingy, whereas members of Groups G and J were characterized by the traits moody, intelligent, and forgetful. Within each pairing, the status of the three traits was counterbalanced across participants, with each trait sometimes serving as the perfect predictor for one group, sometimes serving as the perfect predictor for the other group, and sometimes serving as the imperfect predictor. Following each target presentation and response, participants received feedback informing them of the correct group membership of the target.

The learning task consisted of 15 blocks, with eight target trials in each block. Within each block, three targets were members of each common group (Groups F and J), and one target was a member of each rare group (Groups G and H). Thus, 75% of the targets were members of the majority groups. Participants were given two 30-s breaks, one after the fifth block and one after the 10th block.

After the learning task was complete, participants were presented with new targets possessing novel combinations of traits and were asked to categorize the targets into the four groups without feedback. For each pair of groups (F + H, G + J), two test targets possessed the perfect predictor for the common group (PC) alone, two test targets possessed the perfect predictor for the rare group (PR) alone, two test targets possessed the imperfect predictor (I) alone, four test targets possessed both of the perfect predictors (PC + PR), and four test targets possessed all three traits (PC + PR + I).

Results

Learning phase. The results from 3 participants were eliminated because of poor performance on the final trial (<60%), indicating a failure to learn the group structures. The overall mean for performance on the final trial was 90.19% correct.

Because they appear more frequently, the common groups should be learned before the rare groups. As can be seen in Figure 2, this was the case. To quantify this effect, the difference in accuracy between the common and rare groups was regressed on block number (1–15). This analysis demonstrated a significant negative relationship ($\beta = -.81$, $p < .001$). Thus, the difference in performance between the common and rare groups decreases as learning proceeds.

Test phase. Results from the test phase confirmed that participants learned well the perfect predictors (see Table 1). When presented alone, the perfect predictors for the common groups (PC) were assigned to the common groups at a rate significantly

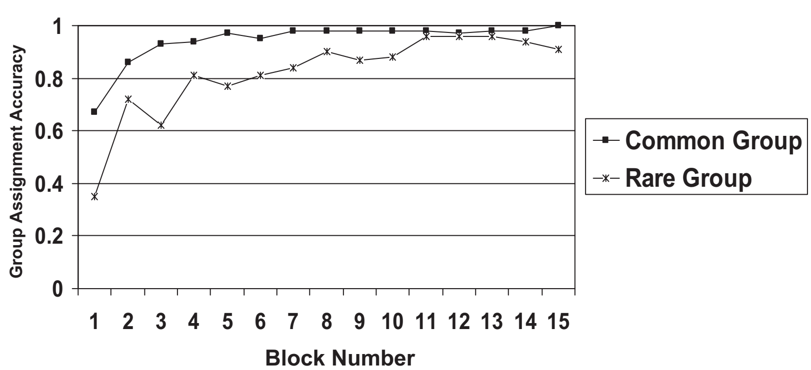

Figure 2. Training data, Experiment 1.

Table 1
Results From Test Trials of Experiment 1

Attributes	Group chosen C	R	CO	RO
PC	.779	.066	.066	.088
PR	.044	.750	.110	.096
I	.588	.272	.044	.096
PC + PR	.346	.507	.029	.118
PC + PR + I	.434	.397	.022	.147

Note. C = common; R = rare; CO = the other common group; RO = the other rare group; PC = perfect predictor of the common group; PR = perfect predictor of the rare group; I = imperfect predictor of the two groups.

greater than chance (78%), $\chi^2(1, N = 35) = 81.82$, $p < .00001$. Similarly, the perfect predictors (PR) of the rare groups were assigned to the rare groups quite accurately (75%), $\chi^2(1, N = 35) = 85.33$, $p < .00001$. Imperfect predictors (I) presented alone were more likely to be assigned to the common (58%) than the rare (27%) group, $\chi^2(1, N = 35) = 15.80$, $p < .0001$.

The results showed a strong inverse base-rate effect for test cases possessing both perfect predictors (PC + PR). Contrary to the base rates, participants were more likely to assign these targets to the rare (51%) than the common (35%) group, $\chi^2(1, N = 35) = 4.17$, $p < .05$.

Finally, when all three traits (PC + PR + I) were present in test cases, assignments were equally likely to be made to the common (43%) and rare (40%) groups, $\chi^2(1, N = 35) = 0.22$, $p = .66$. Thus, the addition of the imperfect predictor was sufficient to counteract the inverse base-rate effect but failed to produce performance in accordance with the base rates.

Discussion

The results of Experiment 1 demonstrated the inverse base-rate effect in the domain of stereotype formation. In so doing, they revealed two novel and important insights about stereotyping. First, the test results with only the imperfect predictor (I) showed that, if a trait is highly descriptive of both a majority and a minority group, as might be true with generally good traits such as *nice*, it is likely to be associated primarily with the majority group. This may be seen as another instantiation of illusory correlation, in that a trait that is equally descriptive of two groups is, nonetheless, associated primarily with one of the groups. Second, the test results with the two perfect predictors (PC + PR) showed that minority group attributes are more predictive of group membership than are majority group attributes. Thus, minority group members are more strongly associated with their traits than are majority group members with theirs. When minority- and majority-stereotypic traits were present in the same individual, the minority trait was given greater weight in categorizing the target. To our knowledge, this is the first demonstration of such an effect. This may help to explain why the most salient stereotypes often are those attributed to minority groups.

Past research (e.g., Kruschke, 1996, 2001; Kruschke et al., 2005) has shown that AT provides the best account of the inverse base-rate effect. According to AT, because common category exemplars (majority group members) are encountered more frequently than rare category exemplars (minority group members), traits that describe both categories (imperfect predictors, I) are associated with the common category first. These traits do not become associated with the rare category even though they are equally descriptive of that category. In turn, rare categories are associated more strongly with their typical attributes than are common categories because of attention shifting when learning about rare category exemplars. Specifically, to distinguish the rare category from the common category, attention is directed away from imperfect predictors (I) when learning about rare category exemplars and toward the perfect predictors of those categories (PR). In contrast, when learning about common category exemplars, attention is divided between imperfect predictors (I) and perfect predictors (PC). As a result, the rare category is associated more strongly with its perfect predictor than is the common category.

The purpose of the attention-shifting process in AT is to preserve learning about the common category while simultaneously facilitating learning of the rare category. One outcome is that there is a stronger relationship between the rare group and its perfect predictor (PR) than between the common group and its perfect predictor (PC). This effect mirrors the category accentuation effects observed in the social psychological literature on stereotype formation. Specifically, that research has shown that participants pay more attention to and weigh more heavily the features of group members that differentiate them from another group (Corneille & Judd, 1999; Eiser, 1971; Krueger & Rothbart, 1990; Krueger et al., 1989; Tajfel & Wilkes, 1963). All of these processes share the common outcome of exaggerating real differences between categories. At the same time, the results for the imperfect predictor (I) demonstrated the formation of differential group impressions on an attribute for which no differences exist. Specifically, participants associated the imperfect predictor (I) more strongly with the common than the rare group even though it was perfectly associated with both groups. In Experiments 2 and 3, we examined AT's ability to account for the perception of group differences that are not real.

Experiment 2

Overview and Predictions

As described above, AT explains the distinctiveness-based illusory correlation by positing that, after first forming a positive impression of Group A, attention shifts to the negative behaviors of Group B, to distinguish it from Group A. Thus, according to this explanation, Group A is more strongly associated with positive attributes than Group B, and Group B is more strongly associated with negative attributes than Group A, thereby producing the typical illusory correlation findings.

However, it is not possible to provide a clear test of the AT account in the standard illusory correlation paradigm because there are not two distinct dimensions. Rather, the group descriptions differ in terms of a single global evaluative dimension (positive–negative). The standard illusory correlation results show that Group A is judged more favorably along this evaluative dimension than is Group B but cannot show that Groups A and B are associated differentially with different dimensions. That is, it is impossible to identify independent positive and negative impressions of the two groups. Thus, one cannot distinguish whether Group A is more positive than Group B, Group A is less negative than Group B, or Group A is both more positive and less negative than Group B. Because drawing these distinctions is critical for testing an AT account of the effect, Experiment 2 used a modified version of the standard illusory correlation paradigm.

As just described, in the standard illusory correlation experiment, the common and rare group attributes are positive and negative evaluative descriptors. For example, in the original demonstration (Hamilton & Gifford, 1976), the majority group (Group A) consisted of 18 positive and 8 negative members, whereas the minority group (Group B) consisted of 9 positive and 4 negative members. In our modified version of this design, the common and rare attributes were independent trait dimensions. Thus, for example, for some participants, Group A consisted of 16 intelligent and 8 friendly members, whereas Group B consisted of 8 intelligent and 4 friendly members. This design maintains the essential features of the illusory correlation paradigm: The majority group is twice the size of the minority group, one trait is more frequent than the other, and the ratios of the two traits are identical both between and within the two groups. However, in this case, we are able to examine differences in perceptions of the two groups independently for the common and rare trait attributes, permitting tests of the AT explanation for the illusory correlation. Specifically, in the above example, AT predicts that Group A will be judged as more intelligent than Group B and that Group B will be judged as more friendly than Group A.[3]

Method

Participants. For their participation, 55 students at Northwestern University were given partial course credit in an introductory psychology course. Participants were run in sessions of 1–4 people.

Materials and procedure. Participants were asked to engage in an impression formation task. They were told that they would be learning about two groups of people, Group A and Group B, and that their task was to form impressions of the two groups. Twenty-four descriptions of members of Group A and 12 descriptions of members of Group B were presented for 2,500 ms each. All descriptions had the form "Name (e.g., Dave), a member of Group (A or B), is Trait (Friendly or Intelligent)." Each participant had a common trait that described 16 members of Group A and 8 members of Group B and a rare trait that described 8 members of Group A and 4 members of Group B. The two traits were friendly and intelligent. The assignment of these traits to common versus rare status was counterbalanced across subjects. For half of the participants, all members of both groups also were described by a second trait, lazy. This second trait occupied the role of the imperfect predictor (I) in the inverse base-rate design (as in Experiment 1). This exploratory variable was included to examine whether the inclusion of additional, nondiagnostic information about each group member would influence the results.

After the impression formation task was complete, participants engaged in a brief filler task, in which they were asked to solve a series of word puzzles. The purpose of the filler task was to clear short-term memory of the group information.

Following the filler task, participants completed the dependent measures. The first measure was a group member assignment task, in which participants were presented with the names and traits of the group members encountered during the impression formation phase and were asked to assign the targets to their appropriate group. These test items always took the form "Name (e.g., Dave), a member of ________ is FRIENDLY. Is Dave a member of Group A or Group B?" The accuracy of these judgments served as a dependent measure, as did the overall extent of assignments of common and rare trait targets to the two groups. The second measure was a trait-rating task, in which participants were asked to rate Groups A and B in terms of both friendliness and intelligence on scales from 1 to 10. Finally, participants were presented with a blank 2 $\times$ 2 table representing Groups A and B and the traits friendly and intelligent. They were told that they had been presented with 36 group members during the impression formation task and were asked to divide those 36 members into the four categories (Group A friendly, Group A intelligent, Group B friendly, and Group B intelligent). The proportion of common/rare trait targets assigned to each group served as the dependent measure.

Results

Trait ratings. AT predicts that Group A would be rated higher than Group B on the common trait, whereas Group B would be rated higher than Group A on the rare trait. Thus, we predicted a crossover interaction between group and trait. A 2 (common trait: friendly vs. intelligent) $\times$ 2 (presence vs. absence of second trait during impression formation) $\times$ 2 (group: A vs. B) $\times$ 2 (trait: common vs. rare) analysis of variance (ANOVA) with repeated measures on the last two variables tested these hypotheses. This analysis revealed a main effect for trait, such that ratings on the common trait were higher than ratings on the rare trait, $F(1, 51) = 24.48$, $p < .001$. Supporting the predictions, this analysis also demonstrated a marginally significant interaction between group and trait, $F(1, 51) = 3.77$, $p < .06$ (see Figure 3). The difference in impressions of Group A ($M = 7.44$) and Group B ($M = 6.91$) on the common trait was marginally reliable, $F(1, 102) = 2.25$, $p < .07$, one-tailed, and the difference in impressions of Group A ($M = 5.80$) and Group B ($M = 6.44$) on the rare trait was reliable, $F(1, 102) = 3.56$, $p = .03$, one-tailed.[4]

Group member estimates. A second way to test the predictions of AT is to examine participants' estimates of the numbers of targets possessing common and rare traits that belong to each group. For common traits, AT predicts that participants should correctly estimate that a greater number of targets possessing the

[3] Technically, there are real differences between Group A and Group B: Group A has more members that possess the common and the rare traits than Group B. However, the extent to which each trait characterizes the members of the two groups is identical (2:1 in favor of Group A). Note that the standard illusory correlation paradigm has the exact same structure: Group A is characterized by a greater number of positive and negative behaviors than Group B, but the ratios of positive and negative behaviors describing the two groups are identical. In the current design, an illusory correlation would be demonstrated if participants perceive the two traits to differentially characterize the two groups. In the standard design, it is impossible to examine positive and negative impressions separately.

[4] The only other effect to emerge from this analysis was a Common Trait (Friendly vs. Intelligent) $\times$ Presence Versus Absence of Second Trait $\times$ Group interaction. This interaction is difficult to interpret and does not moderate the predicted two-way interaction.

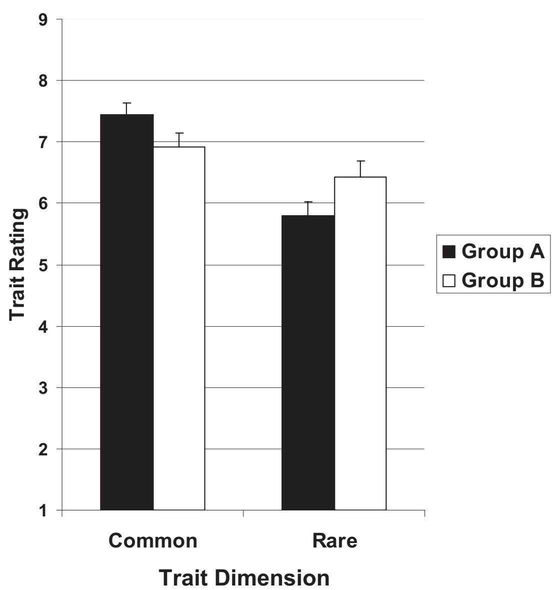

Figure 3. Trait ratings of Groups A and B, Experiment 2. Error bars indicate standard errors.

trait belong to Group A than Group B. However, this tendency should be reduced for rare traits, with a relatively greater number of targets assigned to Group B. This hypothesis was tested with a chi-square analysis on the total numbers of common and rare targets assigned to Group A and Group B, summed across participants. Consistent with this hypothesis, participants assigned proportionately more rare trait targets (M = .51) than common trait targets (M = .42) to Group B, despite the fact that both traits had the same base rates (.33) in Group B, $\chi^2(1, N = 55) = 19.29, p <$ .001. Thus, despite an actual 2:1 ratio of rare trait targets belonging to Group A versus Group B, participants were slightly more likely to assign a rare trait target to Group B.

Target group assignments. The third dependent measure was taken from the group assignment task, for which participants attempted to assign name–trait target pairs to their correct group. For each participant, the proportion of targets possessing rare traits and common traits assigned to Group B (vs. Group A) was determined. Consistent with the group member estimates data described above, a 2 (common trait: friendly vs. intelligent) × 2 (presence vs. absence of second trait during impression formation) × 2 (trait: common vs. rare) ANOVA with repeated measures on the last variable demonstrated that a greater proportion of rare (M = .52) than common (M = .42) trait targets was assigned to Group B, $F(1, 51) = 4.10$, $p < .05$. No other effects were reliable.

The accuracy of target assignments also was analyzed. According to AT, participants learn that the common trait describes Group A and that the rare trait describes Group B. Moreover, these effects are proposed to be driven by an attentional bias, such that, for Group A, common trait targets are attended to more carefully than rare trait targets, whereas for Group B, rare trait targets are attended to more carefully than are common trait targets. As such, AT predicts that, for Group A, assignments should be more accurate for common than rare trait targets, whereas, for Group B, assignments should be more accurate for rare than common trait targets. For each participant, the proportions of common and rare traits assigned accurately (vs. inaccurately) to Groups A and B were determined. These proportions were entered into a 2 (common trait: friendly vs. intelligent) × 2 (presence vs. absence of second trait during impression formation) × 2 (group: A vs. B) × 2 (trait: common vs. rare) ANOVA with repeated measures on the last two variables. This analysis produced a reliable interaction between the group and trait variables, $F(1, 51) = 4.23$, $p < .05$ (see Figure 4). As predicted by AT, for Group B, assignments of rare trait targets (M = .56) were more accurate than assignments of common trait targets (M = .44), $F(1, 102) = 3.98$, $p < .05$, one-tailed. In contrast, for Group A, assignments of common trait targets (M = .60) were more accurate than assignments of rare trait targets (M = .50), $F(1, 102) = 2.94$, $p < .10$, one-tailed.[5] However, although this finding is consistent with an attentional account, it may also have been produced by simple guessing in accordance with participants' base-rate judgments of group–trait associations (i.e., 58% of common trait targets for Group A; 52% of rare trait targets for Group B). We returned to this matter in Experiment 5.

Discussion

The results of Experiment 2 showed that AT can account for the formation of distinctiveness-based illusory correlations. Consistent with the predictions of AT, participants believed that the majority group (Group A) possessed the common trait to a greater extent than did the minority group (Group B) and that the minority group possessed the rare trait to a greater extent than did the majority group. This finding was evident on trait ratings of the groups, estimates of the numbers of group members possessing each trait, and assignments of target individuals to the two groups.

To our knowledge, this is the first demonstration of such illusory correlation effects with independent dimensions representing the common and rare group attributes. We do not believe that any other account of the illusory correlation can accommodate these findings. Explanations that focus on a single dimension of evaluation cannot explain how the two groups would each be associated with a different attribute. For example, the original distinctiveness-based account (Hamilton & Gifford, 1976) posits only that Group B will be seen as possessing the rare attribute more so than Group A. It cannot explain why Group A would be seen to possess the common trait to a greater extent than Group B. Obversely, Rothbart's (1981) account proposes only that Group A will be seen as possessing the common trait to a greater extent than Group B but not that Group B will possess the rare trait to a greater extent than Group A. The three other prominent accounts by Smith (1991), Fiedler (1991), and McGarty and colleagues (e.g., McGarty et al., 1993; for a review, see McGarty & De la Haye, 1997) all explain the illusory correlation by positing that perceivers are attentive to the differences in the raw numbers of common and rare attributes describing each group. These explanations can account for illusory correlation when the common and rare attributes all come from the same dimension (e.g., positive and negative behaviors along a single

[5] The only other effect to emerge from this analysis was a Common Trait (Friendly vs. Intelligent) × Presence Versus Absence of Second Trait × Trait (Common vs. Rare) interaction. This interaction is difficult to interpret and does not moderate the predicted two-way interaction.

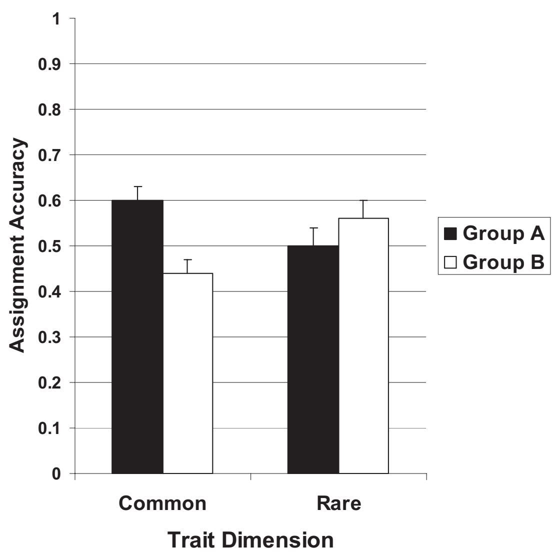

Figure 4. Accuracy of target assignments to Groups A and B, Experiment 2. Error bars indicate standard errors.

evaluative dimension). However, they do not seem well positioned to explain how two distinct dimensions would become differentially associated with the two groups. In this case, each of these theories would need to argue that perceivers are calculating the differences in the extent to which the two independent attributes are descriptive of the two groups. Specifically, in Experiment 2, they would need to argue that perceivers are judging the friendliness and intelligence of the two groups by mentally calculating the differences in numbers of friendly and intelligent members belonging to each group. It is possible that perceivers would perform such calculations, but it does not seem very likely. Because these two traits are unrelated, differences between them are meaningless to the group rating on either trait.[6]

Experiment 3

Overview and Predictions

Experiments 1 and 2 demonstrated the applicability of the AT approach to stereotype formation both when real group differences exist (category accentuation via the inverse base-rate effect) and when they are absent (illusory correlation). It has been proposed here that the AT model can account for both effects as attention shifts away from the majority group and toward the minority group after attributes are first associated with the majority group. However, to bolster the argument that the same mechanism is indeed responsible for both types of effects, it was necessary to demonstrate both phenomena simultaneously within a single procedure.

Experiment 3 consisted of a combination of the paradigms from the first two experiments, in which the characteristics of groups with real differences between them and groups with no real differences between them were learned simultaneously. We predicted that category accentuation and illusory correlation effects would be produced at the same time within this single procedure.

Method

Participants. For their participation, 76 students at Indiana University Bloomington were given partial course credit in an introductory psychology course. Participants were run in sessions of 1–8 people.

Materials and procedure. Participants were asked to engage in a categorization task. They were told they would be learning about four groups: Groups F, G, H, and J.

The procedure essentially combined those employed in the first two experiments, with illusory correlation and inverse base-rate effect trials interspersed throughout the experiment. In the first portion (a total of 72 trials), participants engaged in a training procedure similar to that in the first experiment, in which a name was presented, along with a trait (for illusory correlation trials) or a pair of traits (for inverse base-rate effect trials), followed by a categorization question ("This person belongs to which group? [Press F, G, H, or J]"). On each training trial, participants were provided feedback as to whether their decision had been the correct one.

The inverse base-rate trials (a total of 36) involved making decisions about targets who were members of Groups F or G (a majority group [Ci] and a minority group [Ri], with 27 and 9 members, respectively). This pair of groups was characterized by a real group difference; the two groups shared a trait (honest), and each group was perfectly predicted by a distinct trait (Groups F and G by nosey [PC] and unfair [PR], respectively). On each training trial, subjects were presented with a proper name and a pair of traits—both the imperfect predictor (I) and the perfect predictor (PC or PR for the common and rare groups, respectively)—and were asked to make a decision among all four groups (Groups F, G, H, or J) as to the group to which the person belonged.

The illusory correlation trials (a total of 36) involved making decisions about targets who were actually members of Groups H or J (the majority group [Cx] and minority group [Rx], with 24 and 12 members, respectively). These groups were not characterized by any real intergroup difference; both groups had equal proportions of members with the traits crude and jealous (the common trait [A] and the rare trait [B], with 24 and 12 members, respectively), such that both groups had twice as many members described as crude as described as jealous. Thus, the majority group (H) had 16 crude members and 8 jealous members. The minority group (J) had 8 crude members and 4 jealous members. On each

[6] It should be noted that some of the features of Experiment 2 and of the illusory correlation component of Experiment 3 deviate from the typical illusory correlation procedure. For example, whereas the current research (other than Experiment 5) presented group members paired with trait descriptors, illusory correlation studies typically present descriptions of group members' trait-relevant behaviors (which require more extensive inference processes). Also, in Experiments 2, 3, and 5, participants were instructed to form impressions of the two groups. Previous research has shown that such instructions (as compared with memorization instructions) may reduce or eliminate the illusory correlation effect (e.g., McConnell, Leibold, & Sherman, 1997; McConnell, Sherman, & Hamilton, 1994; Pryor, 1986). Nevertheless, the procedures in the current studies did produce standard illusory correlation effects. Moreover, despite the procedural variations, the fundamental predictions of the different models of illusory correlation remain unchanged, and AT remains the only model that can account for the observed findings.

training trial, subjects were presented with a proper name and a trait—either the common trait or the rare trait—and were asked to make a decision among all four groups (Groups F, G, H, or J) as to the group to which the person belonged.

After the learning phase was complete, participants provided three types of dependent measures, with the order of measures presented in randomized order: target classification, group trait ratings, and group likeability ratings.

On target classification trials, participants engaged in a similar task to that employed in the learning phase. On the screen appeared a proper name, a trait (or set of traits), and a categorization question ("This person belongs to which group? [Press F, G, H, or J]"). However, no feedback was provided for these trials. To test for the inverse base-rate effect, participants made classifications of individuals with the following cue combinations: the shared trait (I), the shared trait paired with the common trait (I + PC), the shared trait paired with the rare trait (I + PR), the common trait paired with the rare trait (PC + PR), and all three traits presented together (I + PC + PR). To test for the illusory correlation effect, participants made classifications of individuals based on the presence of either the common trait (A) or the rare trait (B).

Participants also provided group trait ratings for each group for all traits that had been presented (honest, nosey, unfair, crude, and jealous) on a scale ranging from 1 (*not at all*) to 9 (*very much*).

A third dependent measure had participants rate group likeability for all four groups. Participants rated the likeability of each group ("How likeable is this group?") on a scale ranging from 1 (*not likeable*) to 9 (*very likeable*).

Results

Participants were scored for their level of accuracy on the last training block. This led to the elimination of 18 participants with levels of accuracy below 0.6, leaving 55 subjects to be included in subsequent analyses.

Group classification. The percentages of times each group was chosen given each trait (or set of traits) are presented in Table 2.

Results clearly indicated the presence of an inverse base-rate effect. Corroborating our results from Experiment 1, the shared trait was more strongly associated with the majority group than the minority group. Participants classified the shared trait (I) presented alone as a member of Group F (the majority) 75% of the time and as a member of Group G (the minority) 19% of the time. This preference for the majority group when the shared trait was presented was reliably different from chance, $\chi^2(1, N = 55) = 36.13$, $p < .001$. Most importantly, when the perfect predictors for the two groups were presented together (PC + PR), the preference was found to reverse, with Group G (the minority) selected 70% of the time and Group F (the majority) selected 22% of the time. This preference was reliably different from chance, $\chi^2(1, N = 55) = 27.81$, $p < .001$, and strongly demonstrated the inverse base-rate effect. As in the results of Experiment 1, a stronger relationship emerged between the rare predictor and the minority group than between the common predictor and the majority group.

Providing additional evidence for the particularly strong association between minority groups and their traits, when the shared trait was presented along with the predictors for both groups (I + PC + PR), participants favored a minority classification. Participants classified the targets in these trials as members of Group G (the minority) 59% of the time and as members of Group F (the majority) 36% of the time. This preference differed reliably from chance, $\chi^2(1, N = 55) = 5.95$, $p = .015$. This provided strong support for the dominance of the minority group–minority trait (PR) association, in that this association overrode the associations between the majority group and its respective traits (I and PC).

Results also provided strong support for an illusory correlation effect. When presented with Trait A (common), participants tended to choose Group H (majority; 68%) over Group J (minority; 28%). However, when presented with Trait B (rare), participants tended to choose Group J (57%) over Group H (36%). This difference in preferences was reliable, $\chi^2(1, N = 55) = 21.51$, $p < .001$. Thus, even though there were twice as many Group H members as Group J members who had Trait B, participants overwhelmingly said that new members with Trait B were members of Group J, the minority group.[7]

Group trait ratings. Group trait ratings provided additional support for the inverse base-rate effect (see Figure 5). A repeated measures ANOVA on the group (minority vs. majority) and trait (common vs. rare) variables demonstrated a reliable main effect for group, such that members of Group G (the minority) received relatively stronger trait ratings ($M = 5.84$) than members of Group F (the majority group; $M = 4.95$), $F(1, 54) = 11.58$, $p < .001$. This effect was moderated by a significant interaction between group and trait, $F(1, 54) = 107.40$, $p < .001$. Whereas Group F (the majority) was rated as more PC-ish (the common trait; $M = 6.80$) than Group G (the minority; $M = 4.21$), Group G (the minority group) was rated as more PR-ish (the rare trait; $M = 7.47$) than Group F ($M = 3.09$). No other effects were obtained.

Table 2
Percentage of Group Classifications From Experiment 3

	Group chosen (%)			
Attributes	F	G	H	J
I + PC	94	5	1	1
I + PR	12	85	1	3
I	75	19	5	2
PC + PR	22	70	3	5
I + PC + PR	36	59	0	5
A	1	3	68	31
B	3	4	36	57

Note. I = imperfect predictor of the two groups; PC = perfect predictor of the common group; PR = perfect predictor of the rare group; A = common trait; B = rare trait.

[7] To maintain the consistency of the procedures, participants also were presented with targets possessing both the common trait (A) and the rare trait (B). On these trials, targets were more likely to be assigned to the majority group (65%) than the minority group (31%). Thus, although the rare trait was more strongly associated with the minority group than the majority group, this association was not stronger than the association between the majority group and the common trait. Given that the majority category was objectively associated with both traits (i.e., twice as many members of the majority group as of the minority group were described with each trait), it would be quite surprising to find that a target possessing both traits would be assigned to the minority group.

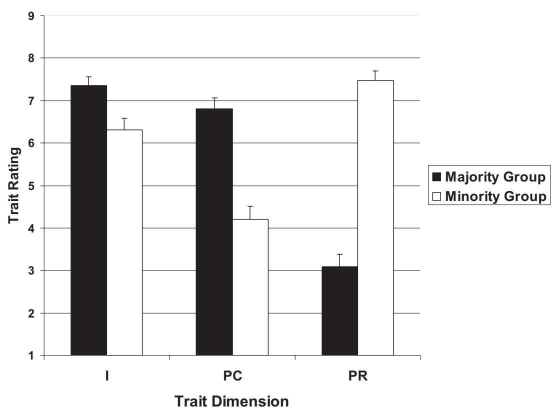

Figure 5. Trait ratings of Majority and Minority Groups, Experiment 3: inverse base-rate effect. Error bars indicate standard errors. I = shared trait; PC = common trait; PR = rare trait.

As anticipated by AT, Group F (the majority) was rated as more I-ish than Group G (the minority), $t(54) = 3.51$, $p = .001$. In other words, the majority group was more strongly associated with the shared trait than was the minority group.

In accordance with the unique predictions of AT, results indicated that the association between the minority group and the rare trait was greater than the association between the majority group and the common trait. This differential association can be demonstrated by comparing the relative degrees of association, conceptualized as *minority predictor trait bias* and *majority predictor trait bias.* The first of these (minority predictor trait bias) was calculated by subtracting the rating for the majority group on the rare trait (PR) from the rating for the minority group on that trait. The second of these (majority predictor trait bias) was calculated by subtracting the rating for the minority group on the common trait (PC) from the rating for the majority group on that trait. The difference between these variables indicates the differential strengths of association between the majority group and its predictive trait and the minority group and its predictive trait, respectively. Indeed, a paired-samples t test demonstrated a significantly greater degree of minority predictor trait bias ($M = 4.38$) than majority predictor trait bias ($M = 2.59$), $t(54) = 3.40$, $p = .001$. Thus, the prediction of a stronger association of traits for the minority group was confirmed.

The data also demonstrated an illusory correlation for Traits A (common) and B (rare), with Group H (the majority) rated to be more A-ish ($M = 6.79$) than Group J (the minority; $M = 6.30$) and Group J rated as more B-ish ($M = 6.76$) than Group H ($M = 6.48$; see Figure 6). This interaction was reliable, $F(1, 54) = 5.39$, $p < .05$, and again supported the development of an illusory correlation. The simple effects for the common and rare traits were not reliable. No other effects were obtained. Again, even though the majority group had twice as many members possessing Trait B as the minority group, the minority group was rated significantly higher on this trait.[8]

Group likeability. Group F (the majority group) received higher ratings of likeability ($M = 5.85$) than Group G (the minority group; $M = 5.01$), $t(54) = 3.11$, $p = .003$. This was consistent with the predictions of our attentional account that the shared trait (I) should be most strongly associated with Group F (the majority group). Because both groups were characterized by one negative trait (either nosey or unfair) and the shared positive trait (honest), the likeability ratings acted as a proxy for the degree of association between the shared trait (I) and the two groups.

For the illusory correlation groups (Groups H and J), no difference in likeability was observed. However, no difference was expected given that all members of both groups were associated only with two negative traits (crude and jealous).

Discussion

Whereas the first two experiments demonstrated the ability of AT to explain the phenomena of both the inverse base-rate effect and the illusory correlation, Experiment 3 demonstrated that both can be produced simultaneously within one procedure. Participants learned simultaneously about a pair of groups with a real intergroup difference and a pair of groups with no actual difference (in the extent to which two different traits described the two groups), and through this method, both category accentuation and illusory correlation effects were demonstrated, as predicted by the AT account. The results of group classifications, trait ratings, and judgments of likeability supported the predictions of our AT account of both the inverse base-rate effect and the illusory correlation. Thus, AT provides a parsimonious explanation for stereotype formation whether or not real group differences exist.

For the inverse base-rate effect, classification of a new target characterized by traits PC and PR (the perfect predictors of the majority and minority groups, respectively) strongly favored the minority group. Thus, even though there were 3 times as many majority group–PC exemplars as there were minority group–PR exemplars, a new target that had both the PC and the PR traits was classified as a minority group member. The trait ratings from the inverse base-rate component of Experiment 3 showed a similar result: The association between the minority group and its associated rare trait was found to be stronger than between the majority group and its associated common trait. This result strongly supported the inverse base-rate prediction that the minority group–PR association was stronger than the majority group–PC association.

For the illusory correlation, the trait ratings showed that the majority group was rated higher on the common trait than was the minority group, whereas the minority group was rated higher on the rare trait than was the majority group. Moreover, in the classification task, targets described with only the rare trait were more likely to be assigned to the minority than to the majority group. These findings occurred despite the fact that the majority group was, in fact, described twice as frequently with the rare trait as the minority group. AT is the only account of the illusory correlation that can explain this finding.

[8] An analysis of predictor trait bias for the illusory correlation traits showed that neither the common nor the rare trait was more strongly associated with the majority and minority groups, respectively. Again, given that the majority group was numerically associated with both traits, it would be surprising to find a stronger association between the rare trait and the minority group than between the common trait and the majority group.

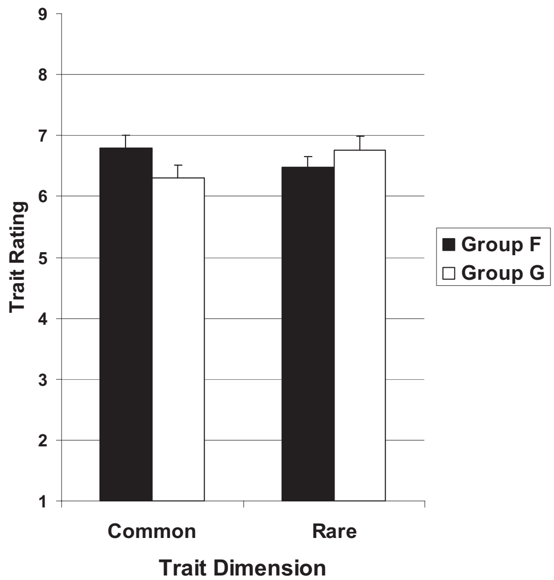

Figure 6. Trait ratings of Groups F and G, Experiment 3: illusory correlation. Error bars indicate standard errors.

Experiment 4

Overview and Hypotheses

The results from Experiments 1 and 3 supported the AT account of category accentuation in the form of the inverse base-rate effect. The fact that a novel target possessing both the perfect descriptors of the majority (PC) and minority (PR) groups was more likely to be classified as a member of the minority group denotes the presence of category accentuation. If the traits describing the two groups were not accentuated, then targets possessing PC and PR would be assigned to the majority group, in accordance with the base rates. The trait ratings from Experiment 3 provided additional support for accentuation in the inverse base-rate effect, in that the minority group was rated more strongly on the PR trait than was the majority group on the PC trait. These ratings were in contrast to the objective extent of pairings between the groups and the traits, in that the groups were each paired perfectly with their traits and there were more instances of the majority group pairing than the minority group pairing. As such, the ratings would seem to reflect accentuation processes.

Nevertheless, we wished to provide further support for the AT account of category accentuation. In contrast to the inverse base-rate designs of Experiments 1 and 3, in Experiment 4, the majority and minority groups were not associated with perfect predictors. Rather, the groups were probabilistically associated with two different traits. Specifically, participants learned about 36 individuals with the common (C) trait (24 from the majority group, 12 from the minority group) and 9 individuals with the rare (R) trait (6 from the minority group, 3 from the majority group). Thus, Trait C was twice as likely to characterize Group A (the majority group) as Group B (the minority group; 24 vs. 12 cases), whereas Trait R was twice as likely to characterize Group B as Group A (6 vs. 3 cases). This design bears some resemblance to the standard illusory correlation design, except that the prevalence of rare traits describing the majority and minority groups were reversed from the usual configuration, with the minority group being described twice as frequently with that trait (see Table 3). As such, the differences in the extent to which the two traits describe the two groups were real, not illusory.

Although there were real differences between the two groups, the groups were associated equally strongly with their respective traits (i.e., there were twice as many Group A as Group B members with the common trait and twice as many Group B members as Group A members with the rare trait). However, according to AT, participants should form stronger associations between the minority group and R than between the majority group and C. Specifically, in learning about the two groups, perceivers will first come to associate the majority group with C because that is the most frequent group–trait pairing. Subsequently, to distinguish the minority group from the majority group, members of the minority group possessing Trait R receive particularly close attention. The greater attention devoted to these targets should also lead to a stronger association between the minority group and R than between the majority group and C, despite the pairs of groups and traits being equally associated objectively (a 2:1 ratio).

A second purpose of this experiment was to further place the predictions of AT against those of extant models of illusory correlation. Even though the differences between the groups in Experiment 4 were not illusory, models of illusory correlation would nevertheless seem to make clear predictions about the nature of the impressions formed of the two groups. In contrast to AT, the distinctiveness-based account (Hamilton & Gifford, 1976) predicts a stronger association between the majority group and the rare trait because that is the most distinctive group–trait pairing in this design. Also in contrast to AT, Rothbart's (1981) account would predict that the strongest association would form between the majority group and the common trait because that is the most frequent group–trait pairing. Neither of these models is designed to explain how the two groups can each become associated with a separate trait. The models of illusory correlation proposed by Smith (1991), Fiedler (1991), and McGarty (e.g., McGarty & De la Haye, 1997) also are not designed to explain the differential associations of two independent traits with two groups. However, the logic of Smith's and McGarty's models implies that the majority group should be more strongly associated with the common trait than is the minority group with the rare trait. This is because the difference in the numbers of common and rare traits describing Group A ($24 - 3 = 21$) is greater than the difference in the numbers of rare and common traits describing Group B ($6 - 12 = -6$). Finally, Fiedler's model suggests that, because there are few presentations of rare traits (the pairing of the majority group with the rare trait is the least frequent), estimates of that trait would show the greatest regression to the mean. As such, this model also

Table 3
Stimulus Presentations for Experiment 4

Attributes	Number of pairings: Group A	Group B
Common	24	12
Rare	3	6

predicts that the association of Group A with the common trait would be stronger than the association of Group B with the rare trait. Again, this prediction contrasts with that of AT.

Method

Participants. For their participation, 73 students at Indiana University Bloomington were given partial course credit in an introductory psychology course. Participants were run in sessions of 1–8 people.

Materials and procedure. Participants were asked to engage in an impression formation task. They were told that they would be learning about two groups of people, Group A and Group B, and that their task was to form impressions of the two groups. The information about Groups A and B consisted of descriptions of behaviors that had been pretested to reflect either the trait friendly (e.g., "Bob, a member of Group A, organized a pot luck dinner for his new neighbors") or the trait intelligent (e.g., "Bill, a member of Group B, taught himself to speak three foreign languages fluently"). In the first part of the impression formation task, participants were presented with a single item at a time for 5,000 ms, for a total of 45 trials describing members of the two groups (as described above). The two traits were counterbalanced, such that friendly was the rare trait for some participants and intelligent was the rare trait for other participants.

After viewing the stimuli, participants rated the degree to which the groups possessed the traits friendliness and intelligence on scales from 1 to 10. Participants then engaged in a target classification procedure, in which they were presented with new targets and accompanying behaviors and were instructed to guess the group membership of each target. This task, which involved evaluating four new targets, was meant to assess how participants applied what they had learned about the groups to newly encountered individuals. For one of these targets, the name was presented with a behavior indicative of the rare trait. For another target, the name was presented with a behavior indicative of the common trait. For two additional targets, the names were presented with pairs of behaviors, one indicative of the rare trait and one indicative of the common trait.

Results

Group trait ratings. A 2 (group: majority vs. minority) × 2 (trait: common vs. rare) ANOVA demonstrated a significant main effect for group, such that the minority group was rated more strongly than the majority group, $F(1, 72) = 5.00$, $p = .029$. A significant main effect was also observed for the trait variable, such that the common trait was rated more strongly than the rare trait, $F(1, 72) = 18.21$, $p < .001$. As predicted, these effects were moderated by a significant interaction between group and trait, $F(1, 72) = 37.19$, $p < .001$ (see Figure 7). Specifically, for the common trait, Group A (the majority group; $M = 8.93$) received higher ratings than Group B (the minority group; $M = 8.26$), whereas for the rare trait, Group B ($M = 8.51$) received higher ratings than Group A ($M = 7.33$).

The key prediction for this measure was that the strength of association between the minority group and the rare trait would be stronger than that between the majority group and the common trait. To test this prediction, we created two bias scores for each participant in the same manner as in Experiment 3. The first of these (*minority rare trait bias*) was calculated by subtracting the rating for Group A (the majority group) on the rare trait from the rating for Group B (the minority group) on the rare trait. The second of these (*majority common trait bias*) was calculated by subtracting the rating for Group B on the common trait from the rating of Group A on the common trait. The difference between these variables indicates the differential strengths of association between the majority group and its predictive trait and the minority group and its predictive trait, respectively. A paired-samples t test demonstrated a significantly greater degree of minority rare trait bias ($M = 1.18$) than of majority common trait bias ($M = 0.67$), $t(72) = 2.235$, $p = .029$. This result provides strong support for the AT prediction that the association between a minority group and a rare trait will be stronger than that between a majority group and a common trait, despite equal likelihoods of group-trait pairings in the two cases.

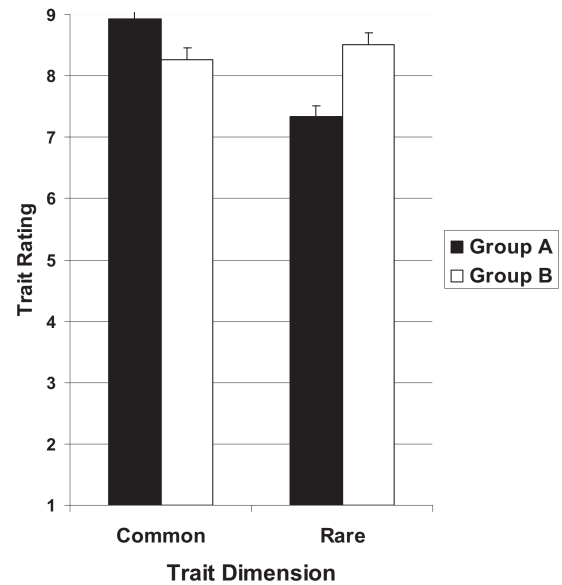

Figure 7. Trait ratings of Groups A and B, Experiment 4. Error bars indicate standard errors.

New target classification. As expected, on the basis of the stimulus configuration, a new target with only the common trait was more likely to be categorized as a member of Group A (the majority group; 67%) than of Group B (the minority group; 33%). A one-tailed binomial analysis of the response frequencies showed this difference to be significant ($N = 73$, $p = .001$). Also as expected, a one-tailed binomial analysis showed that a new target with only the rare trait was more likely to be categorized as a member of Group B (78%) than of Group A (22%;, $N = 73$, $p < .001$). Additionally, mirroring the trait ratings, results indicated that the association between Group B and the rare trait was stronger than the association between Group A and the common trait. In a similar fashion to the trait-rating analysis, we assessed minority rare trait bias (calculated by subtracting the frequency with which targets possessing the rare trait were classified as Group A members from the frequency with which those targets were classified as Group B members) and majority common trait bias (calculated by subtracting the frequency with which targets possessing the common trait were classified as Group B members from the frequency with which those targets were classified as Group A members). A one-tailed binomial test revealed that the degree of Group B rare trait bias was greater than the degree of Group A common trait bias ($N = 66$, $p = .01$). In other words, the degree of preference for classifying a target as a Group B member given the rare trait was greater than the degree of preference for classifying a target as a Group A member given the common trait.

The trials on which the common trait and rare trait were paired for a new target provided another key test of the prediction of greater strength for the Group B–rare trait association compared with the Group A–common trait association. If the rare trait is more strongly associated with Group B than the common trait is with Group A, then when a participant must select a group assignment based on both traits, Group B should be the more likely selection. Indeed, the predictions of AT were supported. Given a target who had both the rare and the common traits, a one-tailed binomial test revealed that target assignments to Group B (58%) were more frequent than target assignments to Group A (42%; $N = 146$, $p = .04$). This occurred despite the fact that selecting Group B was in violation of the base rates of group membership (27 Group A members vs. 18 Group B members).[9]

Discussion

Experiment 4 provided further support for the AT account of category accentuation. Results for both the trait ratings and the group assignments showed that the association between a minority group and a rare trait was stronger than the association between a majority group and a common trait. This was true despite the fact that each group was associated with its respective trait 67% of the time. It is important to note that these results reflect more than a standard accentuation effect, in which real differences between two groups are exaggerated. Rather, specific to an AT account of stereotype formation, the results showed that perceptions of the minority group were accentuated more than were the perceptions of the majority group. We know of no other model of category accentuation that makes this prediction. Furthermore, extant models of illusory correlation, though not specifically designed to account for the real group differences presented in this experiment, do not predict the observed pattern of results.

Experiment 5

Overview and Predictions

In Experiment 5, we sought direct support for the attentional mechanism proposed by AT in the context of the distinctiveness-based illusory correlation. AT argues that perceivers first associate common attributes with the common category and subsequently associate rare attributes with the rare category to differentiate it from the common category. This process is facilitated by an attention-shifting mechanism that causes perceivers to attend most carefully to group–attribute pairs that promote this differentiation (e.g., Kruschke, 1996; Kruschke et al., 2005). Eye-tracking data have provided direct evidence for this mechanism in an inverse base-rate effect study (Kruschke et al., 2005).

In the context of the illusory correlation, AT suggests that, when learning about Group A (the majority group), targets possessing the common trait will be attended to more carefully than targets possessing the rare trait. In contrast, when learning about Group B (the minority group), targets possessing the rare trait will be attended to more carefully than targets possessing the common trait. The target group assignment data from Experiment 2 are consistent with this hypothesis in that, for Group A, assignments were more accurate for common than rare trait targets and, for Group B, assignments were more accurate for rare than common trait targets. However, though this finding is consistent with an attentional account, it may also have been produced by a response bias to assign common trait targets to Group A and rare trait targets to Group B. That is, even if the targets were attended to equally, they may have been assigned differentially to the groups on the basis of judgment-consistent response biases (e.g., because Group A was more intelligent, intelligent targets probably belonged to Group A; Stangor & McMillan, 1992) or on the basis of base-rate consistent guessing (see Experiment 2).

Other past research on the illusory correlation is partly consistent with the attentional hypothesis. Specifically, Stroessner, Hamilton, and Mackie (1992) found that perceivers spent more time reading Group B rare behaviors than any of the other three group–attribute combinations. However, contrary to the predictions of AT, participants spent slightly more time reading rare than common behaviors about Group A. One factor that may have limited the extent of AT-like attentional shifting is the fact that the information about the groups was presented sequentially, a single item at a time. As a result, there was no need to selectively attend to competing items of information within any given trial. In contrast, the attention-shifting component of AT was developed specifically to account for cases in which common and rare attributes directly vie for attention (e.g., Kruschke, 1996; Kruschke et al., 2005). Such situations provide the strongest test of selective attention because the attention directed toward common (rare) attributes comes at the direct expense of attending to rare (common) attributes.

The purpose of Experiment 5 was to test the AT attention-shifting mechanism in an illusory correlation design in which the common and rare attribute information directly competed for attention. We also wanted to test this attentional mechanism in a design in which the common and rare traits denoted distinct dimensions rather than opposite ends of a single evaluative dimension.

Method

Participants. For their participation, 66 students at Indiana University were given partial course credit in an introductory psychology course.

[9] A normative answer to the categorization of a target with both the common trait and the rare trait cannot be assessed. Applying Bayes's theorem (the probability of Group A [or Group B] given both traits) requires additional assumptions about the samples and the sampling procedures. However, the evidence in favor of Group A (24 targets vs. 12 targets for Group B) is based on a much larger sample than the evidence in favor of Group B (6 targets vs. 3 targets for Group A). Thus, the evidence in favor of Group A is more diagnostic, and a categorization of a target with both traits as a minority Group B member would not be predicted.

Materials and procedure. Participants were asked to engage in an impression formation task. They were told that they would be learning about two groups of people, Group A and Group B, and that their task was to form impressions of the two groups. The information about Groups A and B consisted of descriptions of behaviors that had been pretested to reflect either the trait friendly (e.g., "Bob, a member of Group A, organized a pot luck dinner for his new neighbors") or the trait intelligent (e.g., "Bill, a member of Group B, taught himself to speak three foreign languages fluently"). In the first part of the impression formation task, participants were presented with a single item at a time to mirror the standard illusory correlation procedure. Each item was presented for 3,000 ms. Participants learned one, three, or five blocks of information about the groups. Each block consisted of nine items, six describing Group A and three describing Group B. Of the six items describing Group A, four reflected the common trait and two reflected the rare trait. Of the three items describing Group B, two reflected the common trait and one reflected the rare trait. Thus, each block of information was a mini-illusory correlation design, with twice as many items describing Group A as Group B and twice as many common as rare trait behaviors. The assignment of the two traits to common versus rare status was counterbalanced across subjects.

After the presentation of this initial information, participants were told that the impression formation task would continue but that they would now be presented with two items at a time, with one item on the left side of their screens and one item on the right. They were told that, in addition, they also would have to monitor the appearance of an X that would appear on either the left or right side of their screens, and press one of two buttons to indicate the location of the X when it appeared. Four pairs of items included one common trait and one rare trait behavior performed by different members of Group A, and four pairs of items included one common trait and one rare trait behavior performed by different members of Group B. For each pair, the two behaviors were presented for a total of 5 s. Previous research has indicated that this is insufficient time to process both behaviors fully (Sherman, Lee, Bessenoff, & Frost, 1998), thereby placing pressure on participants to selectively attend to one item or the other. During the presentation of each pair of behaviors, the X would appear on the same side of the screen as either the common trait or rare trait behavior. The onset of the X probes was manipulated to occur randomly on either the left or right side of the screen at either 2,750 ms, 3,000 ms, 3,250 ms, or 3,500 ms after the presentation of the behaviors. This variation in the onset of the probes prevented participants from being able to predict precisely when it would appear.

Interest centered on the reaction times to identify the location of the X probe as a function of whether the X appeared on the same side of the screen as a common trait behavior or a rare trait behavior. These latencies measured the extent to which participants were attending to the common or rare trait behavior when the X appeared. The more attention being paid to a particular item, the less time it should take to respond to an X probe that appeared in the same position as that item. The validity of this measure as an indication of attention has been supported by numerous research findings across domains (e.g., Boyer et al., 2006; Bradley, Mogg, White, Groom, & de Bono, 1999; Eberhardt, Goff, Purdie, & Davies, 2004; MacLeod, Mathews, & Tata, 1986; Richeson & Trawalter, 2008; Sherman, Conrey, & Groom, 2004; Sherman, Stroessner, Conrey, & Azam, 2005). We predicted that, when reading about members of the majority group (Group A), participants would identify the location of the X probe more quickly when it appeared on the same side of the screen as common trait behaviors. In contrast, when reading about members of the minority group (Group B), participants should identify the location of the X probe more quickly when it appeared on the same side of the screen as rare trait behaviors.

Results

Reaction times to the X probes were entered into a 2 (common trait: friendly vs. intelligent) × 3 (number of blocks of information prior to X-probe task: one vs. three vs. five) × 2 (group: A vs. B) × 2 (location of X probe: same side as common vs. rare trait) ANOVA with repeated measures on the last two variables. Supporting the predictions, this analysis demonstrated a significant interaction between group and location of the X probe, $F(1, 60) = 3.86$, $p = .05$ (see Figure 8). Response times were marginally faster to X probes appearing in the same position as common trait behaviors ($M = 910$ ms) than rare trait behaviors ($M = 976$ ms) when the behaviors described members of Group A, $F(1, 65) = 2.06$, $p < .08$, one-tailed. In contrast, response times were significantly faster to X probes appearing in the same position as rare trait behaviors ($M = 916$ ms) than common trait behaviors ($M = 993$ ms) when the behaviors described members of Group B, $F(1, 65) = 3.60$, $p < .05$, one-tailed. Furthermore, when X probes appeared in the same position as common trait behaviors, response times were faster when the behaviors referred to a member of Group A ($M = 910$ ms) than when they referred to a member of Group B ($M = 993$ ms), $F(1, 65) = 2.75$, $p = .05$, one-tailed. Finally, when X probes appeared in the same position as rare trait behaviors, response times were faster when the behaviors referred to a member of Group B ($M = 916$ ms) than when they referred to a member of Group A ($M = 976$ ms), $F(1, 65) = 2.70$, $p = .05$, one-tailed.

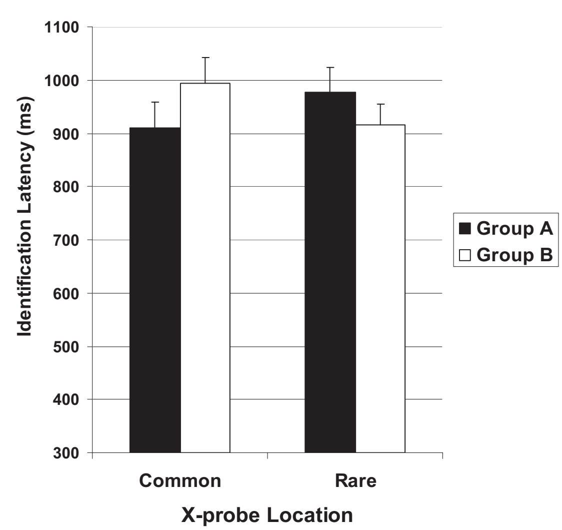

Figure 8. X-probe identification latencies, Experiment 5. Error bars indicate standard errors.

Discussion

The results of Experiment 5 provided further support for an AT account of distinctiveness-based illusory correlation. As predicted by AT, when reading about majority group members, relatively greater attention was paid to common than rare trait behaviors, whereas, when reading about minority group members, relatively greater attention was paid to rare than common trait behaviors. The results also showed that common trait behaviors received more attention when they described majority group members than when they described minority group members, whereas rare trait behaviors received more attention when they described minority group members than when they described majority group members. These results are consistent with the AT proposal that, during category learning, attention is directed toward category members that facilitate discrimination between categories. More specifically, because common traits are first associated with majority groups, minority group exemplars that possess common traits receive relatively little attention compared with minority group exemplars that possess rare traits and majority group exemplars that possess common traits, both of which may be used to differentiate minority groups from majority groups. In contrast, because rare traits may be used to differentiate minority groups from majority groups, members of majority groups possessing rare traits receive less attention than members of majority groups possessing common traits and members of minority groups possessing rare traits.

General Discussion

Stereotype formation may be based on the exaggeration of real group differences (e.g., Tajfel & Wilkes, 1963) or the misperception of group differences that do not exist (e.g., Hamilton & Gifford, 1976). Though related, these phenomena have been explained by different mechanisms. The purpose of the present research was to examine the AT model of categorization (Kruschke, 1996, 2001, 2003) as a basis for integrating these lines of research.

According to AT, the features of majority groups are learned earlier than the features of minority groups. In turn, the features that become associated with a minority are those that most distinguish it from the majority. This latter process is driven by an attention-shifting mechanism that directs attention toward group–attribute pairings that facilitate differentiation of the two groups. Importantly, these processes do not depend on the presence or absence of actual group differences. As long as observers form impressions of one group before the other, then these processes will lead to the formation of distinct stereotypes characterizing the two groups.

The results of five experiments showed that AT provides a valuable framework for integrating research on stereotype formation. Experiment 1 demonstrated the formation of majority and minority stereotypes in the context of an inverse base-rate effect, a case in which both real differences between the groups are exaggerated (for the perfect predictors) and differences are perceived where none exist (for the imperfect predictor). AT has proven to offer the best explanation of this effect (Kruschke, 1996, 2001; Kruschke et al., 2005). Experiment 2 demonstrated that the same AT processes can account for the formation of distinctiveness-based illusory correlations, a context in which there are no real differences in the extent to which different traits describe two groups. Experiment 3 showed that the inverse base-rate effect and the distinctiveness-based illusory correlation could be produced simultaneously within the same procedure, lending support to the possibility that they are produced by a common mechanism. This experiment also provided additional evidence for the operation of category accentuation in the inverse base-rate effect and for the role of AT in producing that accentuation. Experiment 4 provided further support for the AT account of category accentuation and, at the same time, produced findings incompatible with other models of stereotype formation. Finally, Experiment 5 provided direct evidence for the operation of AT's attention-shifting mechanism in stereotype formation.

Altogether, these results show that a common set of processes can explain stereotype formation that occurs on the basis of exaggerated real group differences and on the basis of illusory beliefs that are formed in the absence of real differences. In the former case, the real difference is exaggerated through the AT attention-shifting mechanism, as well as other accentuation processes (e.g., Krueger & Rothbart, 1990; Krueger et al., 1989; Tajfel & Wilkes, 1963). It is possible that the AT attention mechanism may underlie a variety of other accentuation processes involving biased perception, judgment, and memory. In the latter case, though there is no real difference in the extent to which different traits describe the groups, the features of one group are learned prior to the features of the other group because of to their different base rates. This is sufficient to produce the AT attention-shifting process and, presumably, other bases of category differentiation as well (e.g., McGarty & De la Haye, 1997).

Implications for Illusory Correlation

The AT approach also offers an important extension of research on the distinctiveness-based illusory correlation. Up to this point, research on the effect has focused on the extent to which minority and majority groups are perceived along a single evaluative dimension. The typical result is that, when the frequent group descriptors are favorable and the infrequent descriptors are unfavorable, then the minority group is perceived to be less favorable. One difficulty with this approach is that it is often hard to separate the extent to which the minority group is associated with the infrequent descriptor (unfavorable) from the extent to which the majority group is associated with the frequent descriptor (favorable). In contrast, by assigning the frequent and infrequent descriptors to orthogonal trait dimensions (e.g., friendly and intelligent), the present research was able to examine independently the extent to which each dimension was associated with both the minority and majority groups. The results showed that the illusory correlation may be a more complex phenomenon than has been realized. In particular, Experiments 2 and 3 showed that, whereas the minority group is perceived as possessing the rare descriptor to a greater degree, the majority group is perceived as possessing the common descriptor to a greater degree. Thus, it is not simply the case that one group possesses a key dimension to a greater or lesser extent than the other group. Rather, independent stereotypes are developed for each group. Experiment 5 showed that attention is directed differentially toward group–dimension pairings that reinforce these distinct group impressions.

These findings are important because they expand the nature of the illusory correlation effect to be explained and challenge current accounts of the effect. Because past accounts have been designed only to explain differentiation of the minority and majority groups along a single dimension, they have trouble explaining the results of Experiments 2, 3, and 5. For example, accounts designed to explain only why the rare descriptor is associated with the minority group (Hamilton & Gifford, 1976) or why the common descriptor is associated with the majority group (Rothbart, 1981) cannot explain the presence of both effects within a single design. Likewise, accounts that focus on the different numbers of frequent and infrequent descriptors within each group also have a difficult time explaining the results from Experiments 2, 3, and 5 (e.g., Fiedler, 1991; McGarty & De la Haye, 1997; Smith, 1991). According to these accounts, judgments about the central dimension are based on calculations of the extent to which each group is (frequent descriptors) and is not (infrequent descriptors) described as possessing that dimension. When all of the descriptors refer to the same dimension (e.g., an evaluative dimension), such calculations are a relatively simple matter and are sensible (i.e., the difference in the numbers of favorable and unfavorable descriptors for each group). However, it is less apparent how or why perceivers would perform such calculations when the frequent and infrequent attributes refer to distinct and unrelated dimensions (e.g., friendly and intelligent). In this case, these models suggest that perceivers would mentally calculate the difference in the extent to which the two independent attributes describe each group (e.g., friendly minus intelligent). However, it is not clear why perceivers would judge, for example, how friendly each group is by estimating the extent to which each group is relatively more friendly than intelligent.

Extant models of illusory correlation also have a difficult time explaining the results from Experiment 4. Though the group differences in that experiment were not illusory, models of illusory correlation imply clear predictions about the outcomes of the study. The results are not compatible with any of these models of illusory correlation but are explained perfectly by the AT model.

While acknowledging some important differences among the approaches, it is also important to note that, in principle, the AT account is highly compatible with both the distinctiveness account offered by Hamilton and the differentiation account offered by McGarty. Like the differentiation account (e.g., McGarty & De la Haye, 1997), AT argues that a primary motive for perceivers is to form differentiated impressions of the two groups. However, unlike that account, AT does not assume that the differentiation results from an exaggeration of real differences between the groups (i.e., differences in the numbers of frequent and infrequent attributes describing each group). Rather, differentiation results from the fact that the majority group is learned first, and then subsequent processing is oriented toward distinguishing the minority group from the majority group. Like the distinctiveness account (Hamilton & Gifford, 1976), AT argues that special attention is paid to infrequent behaviors performed by members of the minority group. However, unlike that account, AT proposes that the basis for that attention is contextual distinctiveness vis-à-vis the majority group, rather than absolute numerical distinctiveness. Thus, the AT account of illusory correlation integrates the differentiation and distinctiveness approaches in positing that differentiation occurs via a focus on contextually distinct information.

Finally, it is important to reiterate that none of the above is meant to suggest that AT can replace the other accounts of the illusory correlation. Each account has received empirical support, suggesting that the effect is multiply determined. No account, including AT, can account for every point of data. This is certainly true in the case of illusory correlations formed around a single, central dimension.

Further Implications for Stereotype Formation

Beyond the implications for category accentuation and illusory correlation, the AT approach suggests a number of other important novel hypotheses about stereotype formation and impression formation. Some of these hypotheses were tested and supported in the current research, and others await future efforts.

Learning sequence and intergroup comparison. Perhaps the most basic message of AT is that learning sequence matters and that what we learn about a group depends on what we already know about other groups. This observation has a number of important implications for stereotype formation. First, as demonstrated in Experiments 1 and 3, when a trait is highly descriptive of both a majority and a minority group, it is likely to be associated primarily with the majority group. Specifically, in Experiments 1 and 3, even when all members of both the majority and minority groups possessed an imperfect predictor (I), a target possessing only this imperfect predictor (I) was most commonly categorized as a member of the majority group, in line with the base-rate frequencies of the groups. The fact that trait ratings of the I attribute were higher for the majority than the minority group in Experiment 3 indicates that participants not only learned (or guessed) that, probabilistically, a person with Trait I likely belonged to the majority group but also formed an impression (stereotype) that the majority group possessed this trait. These findings suggest that attributes that occur with high frequency in both majority and minority groups are unlikely to become associated with minority groups regardless of how prevalent the attributes may be among those groups. Generally, positively evaluated traits (e.g., nice, friendly, peace-loving) are traits that are likely to occur with high frequency in all groups. The more general point is that the features that are deemed typical of known groups will constrain the types of impressions we may form of unknown groups. Once a trait is taken by one group, the association of the trait with other groups may be inhibited.

This suggests a possible basis for ingroup bias. Because we learn about ingroups prior to learning about outgroups and because the majority of others' behavior is relatively benign, we are likely to form positive impressions of our ingroups. When we subsequently encounter an outgroup, we may be limited in the attributes available for differentiating that group from our own. Most of the attributes that characterize the ingroup will be positive. As such, the best way to differentiate an outgroup from an ingroup may often be by ascribing negative attributes to the outgroup. In a similar manner, perhaps the AT approach can explain the tendency to dehumanize outgroups, especially minority outgroups (Haslam, 2006). Current explanations of dehumanization focus on motivational processes such as delegitimization and moral exclusion. The current AT account would suggest that a generally shared attribute such as humanness is associated with the more frequent group (i.e., one's ingroup), and the traits that distinguish the outgroup from the ingroup are then associated with the outgroup. We do not wish to suggest that this is the most important factor in producing intergroup bias or dehumanization, but it may contribute to the phenomena.

According to this analysis, intergroup comparison is an essential component of stereotype formation. In particular, in considering the attributes of novel groups, the stereotypes of known groups will be brought to mind as a standard. Recent research has supported this prediction (Gawronski, Bodenhausen, & Banse, 2005).

There also are implications for how impressions are formed of individual group members. Consider a case in which a perceiver first meets either a White woman or a Black man and then meets a Black woman. When the first person encountered is a White woman, the feature of the Black woman that distinguishes her from the first target will be her race. In this case, racial stereotypes may play a relatively large role in the impressions formed of the second woman. Knowledge about her race may bias attention, comprehension, memory, and judgment processes toward information that is consistent with whatever racial stereotypes the perceiver holds. When the first person encountered is a Black man, by contrast, the feature of the Black woman that distinguishes her will be her gender. In this case, gender stereotypes may play a larger role in the perceiver's impression of her. Of course, these processes will be maximized to the extent that the perceiver needs to distinguish between the people.

The predictive power of minority stereotypes. Another prediction of AT that was supported in Experiments 1 and 3 is that the attributes that characterize minority groups are given greater weight in categorizing ambiguous targets than are the attributes that are typical of majority groups. Specifically, a target possessing the perfect predictors of both the majority and minority groups (PC + PR) was most commonly categorized as a member of the minority group despite the greater frequency of majority group members. Thus, minority stereotypes had greater predictive power than majority stereotypes. It is commonly observed that the most prominent stereotypes in a given society are those that describe minority rather than majority groups. The current results suggest one possible basis for that observation: To distinguish minority groups from majority groups, particular attention is paid to those attributes that permit differentiation of the minority group. In effect, a stronger link may be formed between the minority group and its typical features than between a majority group and its typical features.

Further implications of this effect for understanding other stereotyping phenomena should be explored in future research. For example, an important aspect of category accentuation that was not examined in the present research is within-group similarity (e.g., Krueger & Clement, 1994; McGarty & Penny, 1988; McGarty & Turner, 1992; Tajfel & Wilkes, 1963). We would expect that the same AT process of attentional accentuation that enhances minority group stereotypes would also lead minority group impressions to be associated with lower levels of perceived intragroup variability than majority group impressions. As another example, the inductive potential in generalizing from individual to individual, from individual to group, and from group to individual may be greater among minority than majority stereotypes. Minority stereotypes also may produce stronger consistency biases in impression formation and memory than majority stereotypes (e.g., Hamilton & Sherman, 1994). Finally, minority stereotypes may be more difficult to alter than majority attributes. Of course, any factor that causes one group's stereotype to be learned subsequent to another group's should produce the same effects. That is, minority status should not be the only basis for such effects.

Conclusion

AT (Kruschke, 1996, 2001; Kruschke et al., 2005) provides a useful framework for integrating different approaches to understanding stereotype formation. This analysis can account for the formation of stereotypes on the basis of both real differences between groups and the perception of differences between groups that are not real. The current research also extends work on the illusory correlation by expanding the nature of the effect to be explained and by challenging existing accounts of the effect. Finally, AT provides a valuable basis for deriving and testing other novel hypotheses about stereotype formation and social perception.

References

Berndsen, M., McGarty, C., Van der Pligt, J., & Spears, R. (2001). Meaning-seeking in the illusory correlation paradigm: The active role of participants in the categorization process. *British Journal of Social Psychology, 40,* 209–233.

Berndsen, M., Van der Pligt, J., Spears, R., & McGarty, C. (1996). Expectation-based and data-based illusory correlation: The effects of confirming versus disconfirming evidence. *European Journal of Social Psychology, 26,* 899–913.

Boyer, M. C., Compas, B. E., Stanger, C., Colletti, R. B., Konik, B. S., Morrow, S. B., & Thomsen, A. H. (2006). Attentional biases to pain and social threat in children with recurrent abdominal pain. *Journal of Pediatric Psychology, 31,* 209–220.

Bradley, B. P., Mogg, K., White, J., Groom, C., & de Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology, 38,* 267–278.

Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorization of multifaceted stimuli. *Journal of Personality and Social Psychology, 77,* 927–941.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87,* 876–893.

Eiser, J. R. (1971). Enhancement of contrast in the absolute judgment of attitude statements. *Journal of Personality and Social Psychology, 17,* 1–10.

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 60,* 24–36.

Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variance. *Journal of Personality and Social Psychology, 63,* 356–367.

Gawronski, B., Bodenhausen, G. V., & Banse, R. (2005). We are, therefore they aren't: Ingroup construal as a standard of comparison for outgroup judgments. *Journal of Experimental Social Psychology, 41,* 515–526.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in intergroup perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12,* 392–407.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology, 39,* 832–845.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer, Jr.,

& T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 2, pp. 1–68). Hillsdale, NJ: Erlbaum.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10,* 252–264.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive stimulation* (pp. 9–33). Coral Gables, FL: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 276–296). New York: Appleton-Century-Crofts.

Krueger, J., & Clement, R. W. (1994). Memory-based judgments about multiple categories: A revision and extension of Tajfel's accentuation theory. *Journal of Personality and Social Psychology, 67,* 35–47.

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology, 59,* 651–663.

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology, 56,* 866–875.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 3–26.

Kruschke, J. K. (2001). The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1385–1400.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12,* 171–175.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 830–845.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95,* 15–20.

McConnell, A. R., Leibold, J. M., & Sherman, S. J. (1997). Within-target illusory correlations and the formation of context-dependent attitudes. *Journal of Personality and Social Psychology, 73,* 675–686.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). On-line and memory-based aspects of individual and group target judgments. *Journal of Personality and Social Psychology, 67,* 173–185.

McGarty, C., & De la Haye, A. (1997). Stereotype formation: Beyond illusory correlation. In R. Spears (Ed.), *The social psychology of stereotyping and group life* (pp. 144–170). Malden, MA: Blackwell.

McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology, 23,* 391–410.

McGarty, C., & Penny, R. E. C. (1988). Categorization, accentuation and social judgement. *British Journal of Social Psychology, 22,* 147–157.

McGarty, C., & Turner, J. C. (1992). The effects of categorization on social judgment. *British Journal of Social Psychology, 31,* 253–268.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Pryor, J. B. (1986). The influence of different encoding sets upon the formation of illusory correlations and group impressions. *Personality and Social Psychology Bulletin, 12,* 216–226.

Queller, S., Schell, T., & Mason, W. (2006). A novel view of between-categories contrast and within-category assimilation. *Journal of Personality and Social Psychology, 91,* 406–422.

Richeson, J. A., & Trawalter, S. (2008). The threat of appearing prejudiced and race-based attentional biases. *Psychological Science, 19,* 98–102.

Rothbart, M. (1981). Memory and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup relations* (pp. 145–181). Hillsdale, NJ: Erlbaum.

Sherman, J. W., Conrey, F. R., & Groom, C. J. (2004). Encoding flexibility revisited: Evidence for enhanced encoding of stereotype-inconsistent information under cognitive load. *Social Cognition, 22,* 214–232.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75,* 589–606.

Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and stereotype maintenance processes: Attention, attribution, and individuation. *Journal of Personality and Social Psychology, 89,* 607–622.

Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 27,* 107–123.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111,* 42–61.

Stroessner, S. J., Hamilton, D. L., & Mackie, D. M. (1992). Affect and stereotyping: The effect of induced mood on distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 62,* 564–576.

Stroessner, S. J., & Plaks, J. E. (2001). Illusory correlation and stereotype formation: Tracing the arc of research over a quarter century. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 247–259). Mahwah, NJ: Erlbaum.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology, 54,* 101–114.

van Rooy, D., van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110,* 536–563.

Received March 27, 2007
Revision received June 27, 2008
Accepted July 11, 2008 ■