

A Novel View of Between-Categories Contrast and Within-Category Assimilation

Sarah Queller
Indiana University

Terry Schell
Rand Corporation

Winter Mason
Indiana University

This research manipulated the portion of a category distribution that is misclassified by the optimal classifier and investigated the impact on assessments of category attributes. Three separate studies manipulated the direction of overlap, the extent of overlap, and the relative base rate of the comparison category. All 3 studies produced large between-categories contrast and within-category assimilation. As expected, these effects were enhanced in conditions in which the optimal classifier misclassified a larger portion of the target category. Study 4 demonstrated that intercategory overlap in the absence of overt classification does not produce contrast and assimilation. Ironically, optimizing categorization accuracy can produce highly inaccurate beliefs about category attributes.

Keywords: stereotype/stereotyping, categorization, stereotype accuracy, contrast, assimilation

A large body of research has focused on how perceivers relegate items that vary along one or more dimensions into discrete classes, or categories. Consequently, the field has made great strides in describing human categorization, and the modeling of the underlying psychological processes continues to stir considerable debate. In the process of focusing closely on how people put *things* into *bins*, cognitive categorization researchers have largely set aside questions related to the importance of categorization. Why do we care about these bins anyway? Presumably, it is because such classifications can simplify our interactions with the myriad items we encounter as we go about our daily lives. For example, when gathering wild mushrooms, the gatherer's ability to tell an amanita from a puffball can be quite important. Once a mushroom is classified, the gatherer knows whether to gather and eat it (puffball) or to steer clear (amanita). Thus, the category membership carries information that guides behavior.

This is easily recognized with social categories as well. A person classified as a subordinate in the social hierarchy might inspire a jovial slap on the back, whereas a person classified as a superior might be approached with more deference. By assigning instances to categories, perceivers are able to infer accurate and useful information about the attributes of that class of stimuli that can guide behavior. Thus, the importance of categories is derived from how attributes are inferred from category membership as much as it is from how category membership is inferred from attributes.

For some time, psychologists have claimed that the act of categorization itself might bias perceivers' beliefs about what a

category is like (Allport, 1954). Tajfel and Wilkes's (1963) early experimental work reinforced these ideas. These researchers labeled the shortest half in a series of lines of graded length with one category label and the longest half in the series with another category label. They demonstrated that the discrepancy between perceivers' estimates of the lengths of the two stimuli nearest the category boundary was exaggerated compared with an unlabeled control condition and a randomly labeled control condition. They expected to find within-category assimilation as well, so that stimuli within the same systematically labeled category would seem more similar to one another than would the same stimuli in either the no label or the randomly labeled control condition. Their data, however, failed to support this latter expectation.

Between-Categories Contrast and Within-Category Assimilation

Such influences had particular import for researchers of social categorization who argued that the act of classifying people into different groups might lead to inaccurate stereotypes. Accordingly, Tajfel and Wilkes's (1963) work has since become a cornerstone of researchers' claims that differences between social groups might be exaggerated. Indeed, the article was cited 46 times between January 2000 and October 2004, mostly in social psychology journals.

More recent research has produced results that are consistent with Tajfel and Wilkes's original hypotheses regarding between-categories contrast (Corneille, Klein, Lambert, & Judd, 2002; Goldstone, 1994) and within-category assimilation (Livingston, Andrews, & Harnad, 1998). Evidence of both effects occurring simultaneously, however, is rare. This is true even though a wide variety of measures have been used to detect changes in perceptions of the category. Stimulus similarity judgments (Livingston et al., 1998), same-different judgments (Goldstone, 1994), and esti-

Sarah Queller and Winter Mason, Department of Psychology, Indiana University; Terry Schell, Rand Corporation, Santa Monica, California.

Correspondence regarding this article should be addressed to Sarah Queller, Department of Psychology, Indiana University, 1101 East 10th Street, Bloomington, IN 47405. E-mail: queller@indiana.edu

mates of the attribute values of specific stimuli (Eiser, 1971; Krueger & Clement, 1994; McGarty & Penny, 1988; Tajfel & Wilkes, 1963) have been relied on to assess how perceivers enhance or reduce differences between stimuli. Other work has focused on biases in judgments about the prototypical category member as a function of intercategory comparisons (Goldstone, 1996; Krueger, Rothbart, & Sriram, 1989). Still other work has considered judgments of stimulus typicality as evidence of contrastive and assimilative biases (Corneille & Judd, 1999).

In some cases, both between-categories contrast and within-category assimilation are present in the same study, but each is found using a different dependent measure (Goldstone, Lippa, & Shiffrin, 2001; Livingston & Andrews, 2005). In these cases, one dependent measure may slant the judgment toward contrastive processing and the other may slant the judgment toward assimilative processing. Although these studies are informative about factors that lead to contrast and assimilation, they have provided little support for the kind of simultaneous between-categories contrast and within-category assimilation to which Tajfel and Wilkes (1963) alluded. Like many of the researchers who followed them, they clearly expected to obtain both effects with the same measure. Thus, despite widespread claims that assessments of category attributes are subject to both between-categories contrast and within-category assimilation, both effects have not been consistently evident in experimental studies.

An Effect in Search of a Mechanism

Although much of the research in this area has been directed toward demonstrating the existence of categorization-driven contrast and assimilation effects, research has also been directed at explicating the mechanisms that drive these effects. Social psychologists have focused on motivational forces. Describing one set of motivations, Pickett and Brewer (2001) argued that people have competing needs to be both similar to others (assimilative needs) and distinct from others (distinctiveness needs). Biased perceptions of within-category homogeneity and between-categories distinctiveness can serve these goals. People are also prone to exaggerate differences between groups if such exaggeration casts the in-group in a more favorable light than the out-group (Mullen, Brown, & Smith, 1992). Although these motivational mechanisms likely play a role in learning about social groups, they cannot explain contrast and assimilation with stimuli such as line lengths, fish, and chick cloacae (Corneille & Judd, 1999; Corneille et al., 2002; Livingston et al., 1998; Tajfel & Wilkes, 1963). Conversely, however, the mechanisms that explain contrast and assimilation in these nonsocial categories are likely to contribute to learning about social categories. Why, then, do contrast and assimilation occur in these less motivated situations?

One possibility is that categorical perception effects drive between-categories contrast and within-category assimilation. Goldstone's (1994) work is informative in this domain. Using same-different judgments, he showed that perceptual discrimination between similar stimuli is increased along a categorization-relevant dimension and is particularly pronounced at the category boundary. Presumably, such enhanced perceptual discrimination could lead to assessments of group characteristics that exaggerate differences between groups. Replicating Tajfel and Wilkes (1963), Goldstone did not find diminished perceptual discrimination of

stimuli within the same category that, if present, would inflate assessments of within-group homogeneity. These types of perceptual effects have been the basis for applying the original Tajfel and Wilkes (1963) research to the issue of biased stereotypes. Yet it seems unlikely that small perceptual effects would lead to large changes in assessments of a category's central tendency or dispersion. To be relevant to the claim that stereotypes of social groups are biased from reality in a way that might matter for behavioral interaction, these perceptual effects would have to substantially impact assessments of group attributes.

The Present Research

The present work suggests a rather different approach to understanding between-categories contrast and within-category assimilation. The approach was based on the idea that optimizing categorization can lead to systematically biased beliefs about group-level attributes. Note that this is a rather alarming contention: It suggests that optimizing the accuracy of categorization can lead to quite inaccurate knowledge about category attributes.

In connecting categorization optimization with biased perceptions of group attributes, we made underlying assumptions in the present work that distinguish it from many prior studies on between-categories contrast and within-category assimilation (Corneille & Judd, 1999; Corneille et al., 2002; Eiser, 1971; Goldstone, 1994; Krueger & Clement, 1994; Livingston et al., 1998; Tajfel & Wilkes, 1963). These assumptions were that categories often contain large numbers of members and that their distributions often overlap along continuous dimensions. Note that these two assumptions are certainly true of the categories that are of interest to stereotyping researchers. African Americans and European Americans, Democrats and Republicans, blue-collar and white-collar workers, and women and men are just a few of many examples of large categories that overlap in terms of physical attributes (e.g., skin color, height), attitudes (e.g., handguns, welfare), and trait attributes (e.g., social sensitivity, intelligence). These assumptions are also true of many nonsocial categories, such as classes of cars (e.g., gas mileage, cargo capacity) and diseases (e.g., blood pressure, temperature, discharge color).

A consequence of these assumptions is that even the optimal classifier cannot achieve 100% categorization accuracy. To illustrate, suppose you know a man who is rather effeminate, and you want to decide whether that man is gay or straight. You know he is either a gay man or an unusually effeminate straight man because past experience has taught you that men this effeminate are more likely to be gay than straight. In the absence of other cues to inform your judgment, you will maximize your chance of being right (i.e., maximize your classification accuracy), if you classify this man as gay. Of course, some straight men actually are this effeminate, so you will be wrong some of the time. Because you do not know which of the men with this level of effeminacy are gay and which are straight, however, you can achieve higher classification accuracy if you always respond "gay" for men with this level of effeminacy. More generally, there is a level of effeminacy at which gay and straight men are equally likely. Given effeminacy as your only cue, you can maximize your classification accuracy if you consistently respond "gay" for effeminacies greater than the point of equal likelihood and consistently respond "straight" for effeminacies less than that point. For a target that has an effemi-

nacy that is equally likely for gay and straight men, your best bet is to guess at sexual orientation.

The optimal classifier will systematically misclassify those stimuli whose attributes have a higher likelihood of being from the opposing category. Because human perceivers become nearly optimal in their classification after training with feedback (Ashby, 1992; Ashby & Gott, 1988), human perceivers also misclassify those stimuli whose characteristics make them more likely to have come from the opposing category. These misclassified stimuli are precisely those stimuli that are most similar to the comparison category.

A simple assumption of the present theorizing is that systematic misclassification of a specific portion of the category distribution might lead to a biased representation of what the category is like. In the previous example, misclassifying the most effeminate straight men as gay men could lead you to underestimate the average effeminacy of straight men. It could also lead you to underestimate how effeminate the most effeminate straight man is. By a similar argument, you would overestimate the average effeminacy of gay men and the effeminacy of the least effeminate gay man.

Established theory suggests why such bias might occur. For example, exemplar theory (Nosofsky, 1986) assumes that the correct category label that is provided during feedback is stored with each exemplar. Consistently responding "Group B" to the Group A stimuli whose attributes make them more likely to come from Group B could lead to storage of those Group A stimuli with the incorrect Group B label. That is, it may be the response label rather than the feedback label that is stored with each exemplar (Nosofsky & Johansen, 2000). In effect, this is tantamount to storage of a distribution of exemplars for each group that is truncated at the intercategory boundary. Storage of truncated distributions would produce both exaggerated differences between category means and exaggerated within-category homogeneity.

In contrast to exemplar theory, Ashby and colleagues (e.g., Ashby & Gott, 1988) have argued that categorization is optimized not by making judgments based on stored exemplars but by determining the decision bound that optimizes accuracy in categorization. Although these researchers have not been explicit about the representational underpinnings of such boundary-driven responding, one possibility is that the response regions of the stimulus space are stored (Ashby & Casale, 2003). In this case, portions of the stimulus space that correspond to the real category distributions will be ignored in the representation of what the category is like. Assuming that the stored response region is used as the basis of making category-central tendency and dispersion estimates, both between-categories contrast and within-category assimilation would increase as the portions of the categories that are misclassified increase.

There are, undoubtedly, other mechanisms that would weaken the associations between optimally misclassified stimuli and their correct category label and/or strengthen the associations between these stimuli and the incorrect category label. The goal of our research was not to pit these mechanisms against each other. Rather, we simply investigated the claim that optimizing classification accuracy can lead to large and predictable shifts in beliefs about category attributes. In doing so, the present research stands to support Allport's (1954) claim that the simple act of categorization can affect beliefs about social groups.

This work contributes to the extant literature by suggesting a previously uninvestigated basis for between-categories contrast and within-category assimilation. Specifically, between-categories contrast and within-category assimilation are both predicted to occur when a prior goal to optimize categorization accuracy has induced systematic misclassification of portions of the category distributions. Consistent with this argument, four studies manipulated factors that altered the optimally misclassified portions of the category distributions and, as expected, produced predictable shifts in between-categories contrast and within-category assimilation.

Study 1

The optimal responder will misclassify those stimuli whose attributes make them more likely to have come from the opposing category. For a nonideal responder, misclassifications may not be restricted to the optimally misclassified stimuli but will nonetheless occur more frequently for those stimuli whose attributes are similar to those of the contrast category. Thus, if systematic misclassification leads to biased assessments of category attributes, human classifiers will shift their estimates of a fixed target category's central tendency as a function of the direction of overlap with the comparison category.

Using a between-subjects design, Study 1 compared a fixed target Group A with a contrast category that was either more intelligent and less friendly (Group B_i) or one that was more friendly and less intelligent (Group B_f; see Figure 1). To maximize A versus B_i classification accuracy, perceivers will misclassify the most intelligent and least friendly Group A members as belonging to Group B. As a result, Group A was expected to be seen as relatively less intelligent and more friendly than it really was. Similarly, to maximize A versus B_f classification accuracy, we

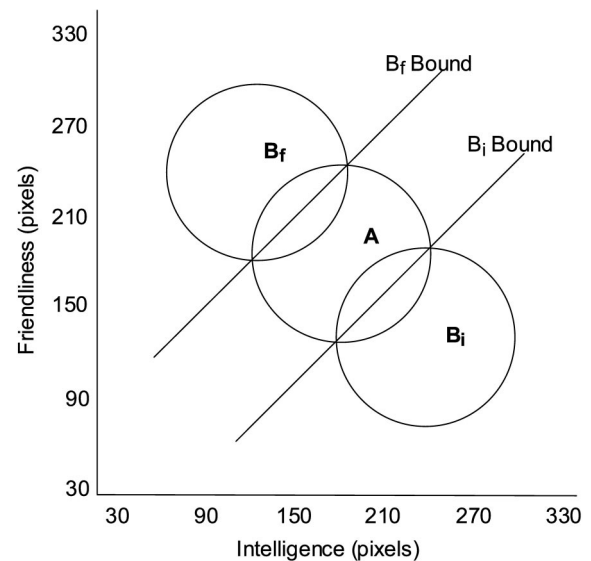


Figure 1. Bivariate normal stimulus distributions for Groups A and B of Study 1. Participants learned about either Group A and a contrast category that was more friendly and less intelligent (B_f) or about Group A and a contrast category that was more intelligent and less friendly (B_i). Studies 2 and 3 use the same Group A stimuli and variations of the Group B_f stimuli as described in the text.

misclassified the most friendly and least intelligent Group A members as belonging to Group B. Consequently, Group A was expected to be seen as less friendly and more intelligent than it really was. Thus, assessments of the target category's attributes were predicted to be systematically contrasted away from the relevant comparison category.

Note, however, that comparison of perceivers' estimates of Group A's attributes to Group A's true distribution parameters does not provide the best test of the hypothesis. Any systematic misjudgment, for example, a tendency to underestimate, would bias the estimates away from the true population parameters. In addition, a spurious main effect, such as a bias to report that people are more intelligent than friendly, would affect comparisons of estimates to actual population parameters. These potential confounds were dealt with by testing for the presence of an interaction between direction of overlap (B_f vs. B_i) and trait (intelligence vs. friendliness). A significant interaction such that a fixed Group A was judged to be relatively more intelligent and less friendly in the B_f condition than in the B_i condition could be taken as evidence of between-categories contrast.

On each trial, participants viewed two bars whose heights depicted the degree of friendliness and intelligence of an individual group member. The participant learned about the groups by deciding whether each person was a member of Group A or Group B, receiving immediate corrective feedback, and continuing on to the next trial. After participants demonstrated the ability to distinguish between the categories at near optimal levels, they made judgments about the fixed target group, Group A.

Method

Design and participants. Study 1 used a 2 (B_f vs. B_i direction of overlap) \times 2 (bar/exemplar based vs. exemplar based/bar counterbalance) \times 2 (intelligence vs. friendliness) design with direction of category overlap as the key variable that was manipulated as a between-subjects factor (B_f vs. B_i). A counterbalance in the order of the dependent measures constituted a second between-subjects factor (bar-exemplar based vs. exemplar based-bar; see the *Procedure* section for details). Information about both intelligence and friendliness was provided for each stimulus, so trait (intelligence vs. friendliness) constituted a third, within-subject factor.

Twenty-six undergraduates enrolled in introductory psychology at the University of California, Santa Barbara, participated in this study and received partial course credit for their participation. Participants were randomly assigned to either the B_f or the B_i condition.

Stimulus materials. The stimuli presented to each participant were displayed on a conventional VGA computer monitor. Each stimulus consisted of two red bars against a black background. The bar on the left was labeled *intelligence* and the bar on the right was labeled *friendliness*. The heights of the bars indicated each stimulus person's level of intelligence and friendliness.

For the training phase of the experiment, 250 stimuli were generated for Group A and 250 stimuli were generated for Group B. For the Group A stimuli, the mean bar height on the two dimensions was equal ($\mu_i = \mu_f = 180$ pixels, see Figure 1). For participants in the B_i condition, Group B was more intelligent ($\mu_i = 235$) and less friendly ($\mu_f = 125$) than Group A. For participants in the B_f condition, Group B was more friendly ($\mu_f = 235$) and less intelligent ($\mu_i = 125$) than Group A. For each group, stimuli were sampled from bivariate normal populations that varied on the dimensions of intelligence and friendliness. The standard deviation on each dimension was equal to 30 pixels, and the covariation between the two dimensions was zero. The sampling of stimuli from these distributions was random with three constraints: (a) no stimuli could be more than 3 standard

deviations from the mean, (b) constants were added to the stimuli from each group so that the sample means were equal to the desired population mean, and (c) a linear decision bound existed that would produce 90 (± 1)% correct categorizations.

For the test phase of the experiment, 121 stimuli were generated. These stimuli were distributed across the stimulus space at equal intervals. Specifically, intelligence and friendliness bar heights ranged from 30 to 330 pixels in 30-pixel increments, and a stimulus was generated for each possible pairing of these intelligence and friendliness bar heights.

The same set of 500 training and 121 test stimuli was shown to all participants. The order of presentation of the stimuli within the training and test phases was randomly determined. However, to reduce between-subjects variability, we presented the stimuli to all participants in the same random order.

Procedure. Participants were placed in individual cubicles and read the following instructions on a computer:

In this part of the experiment, your task will be to learn about the characteristics of people from one of two groups, labeled Group A and Group B. Each person from these groups has been given a personality test that measures both friendliness and intelligence. A graph of these two scores will be displayed for each individual. Your task is to guess the group membership of that person based on these graphs.

In each training trial, the bar graph depicting the intelligence and friendliness of one group member was displayed on the monitor. The participant's categorization judgment was recorded by a keyboard response. Immediately after this response, written feedback was presented above the bar graph. This feedback indicated whether the participant's response was correct or incorrect and provided the correct group label for that individual (e.g., "[in]correct, that was a member of Group A"). The feedback and stimuli remained on the screen for 2 s before the beginning of the next trial.

As soon as 131 of the previous 150 responses were correct (87.3% accuracy), the participant moved on to the test phase. Note that this learning criterion is $\sim 3\%$ short of optimal responding and is well above the optimal accuracy that can be achieved using a single dimension (82.2%). The test phase included both a bar height adjustment measure of group central tendency and an exemplar-based measure of group central tendency. The order of these two tasks was counterbalanced.

For the bar height adjustment measure of central tendency, participants directly estimated the average levels of intelligence and friendliness for Group A. They were presented with a stimulus in which both the intelligence and friendliness bars were initially set to the minimum values. They were told to "adjust the bars so that they represent the scores of the average member of Group A." Participants adjusted the height of a particular bar using the \uparrow and \downarrow keys; they selected whether they were adjusting the intelligence or the friendliness bar by using the \leftarrow or \rightarrow keys. When they were satisfied with their estimation, they pressed the \downarrow key, and the computer recorded the estimates. It is important to note that the starting point for the bar provides a low anchor for participants' assessments of the category average. Thus, it would not be surprising if this anchor produced a bias such that both intelligence and friendliness were underestimated (Mussweiler, 2003; Tversky & Kahneman, 1974). Such an anchoring effect would not, however, jeopardize the power of the study to detect relative differences between the B_i and B_f conditions in estimates of intelligence and friendliness.

For the exemplar-based measure of central tendency, participants viewed each test phase stimulus and pressed one of two keys to indicate whether that stimulus was or was not typical of Group A. By the end of the test phase, we had a set of intelligence–friendliness combinations that participants considered typical of Group A. We then derived an exemplar-based measure of central tendency by computing the average friendliness and intelligence for this set of exemplars. This exemplar-based measure is inherently multidimensional in that participants judge a particular combi-

nation of friendliness and intelligence on each trial, rather than making one judgment on intelligence and a separate judgment on friendliness.

We hoped that the exemplar-based measure would get at perceptions of Group A that were affected as little as possible by comparisons to Group B at the time of judgment. Thus, participants were informed that they would decide whether each test individual was typical of Group A or not, but they were also informed that none of the individuals presented in the test phase would be members of Group B.

Results

Recall that the true Group A mean has equal bar heights for intelligence and friendliness ($\mu_i = \mu_f = 180$). In the B_i condition, in which Group B was more intelligent and less friendly than Group A, maximizing categorization accuracy should lead participants to misclassify the most intelligent and least friendly Group A members. Consequently, these participants' assessments of the Group A average should be less intelligent than friendly. In the B_f condition, in which Group B was more friendly and less intelligent than Group A, maximizing categorization accuracy should lead participants to misclassify the most friendly and least intelligent Group A members. Consequently, these participants' assessments of the Group A average should be less friendly than intelligent. These hypotheses were tested using a 2 (B_f vs. B_i direction of overlap) \times 2 (bar/exemplar based vs. exemplar based/bar counterbalance) \times 2 (intelligence vs. friendliness) analysis of variance (ANOVA). The counterbalance factor had no main effects and was involved in no interactions, and it is not discussed further.

Bar height adjustment task. Five participants were excluded from this analysis because they failed to adjust both bars before continuing with the experiment. The group average judgments for the remaining 21 participants showed the predicted biases (Table 1); the predicted Direction of Overlap \times Trait interaction was significant, $F(1, 17) = 24.52$; $p < .001$.

A simple-effects analysis was conducted to determine whether the predicted bias occurred reliably within each condition. In the B_i condition, participants' central tendency estimates of Group A had intelligence bars that were significantly lower than the corresponding friendliness bars, $F(1, 17) = 5.90$, $p < .05$, $d = 0.54$. In the B_f condition, participants' central tendency estimates of Group A had intelligence bars that were significantly higher than the corresponding friendliness bars, $F(1, 17) = 18.59$, $p < .001$, $d = 2.64$.

There was also a nonpredicted main effect for the trait that was being rated. Averaged across all conditions, participants' central

tendency estimates had intelligence bars that were significantly higher than the corresponding friendliness bars, $M_s = 176.9$ and 150.1 respectively, $F(1, 17) = 12.79$, $p < .01$, $d = 0.74$.

Exemplar-based measure of central tendency. The average intelligence and friendliness bar heights of the exemplars that participants judged to be typical of Group A also showed the predicted biases (Table 1) and the Direction of Overlap \times Trait interaction was significant, $F(1, 22) = 155.2$, $p < .001$.

A simple-effects analysis was conducted to determine whether the bias occurred reliably within each condition. In the B_i condition, the average intelligence bars of exemplars judged typical of Group A were significantly lower than the corresponding friendliness bars, $F(1, 22) = 75.25$, $p < .001$, $d = 3.17$. In the B_f condition, the average intelligence bars of exemplars judged typical of Group A were significantly higher than the corresponding friendliness bars, $F(1, 22) = 79.14$, $p < .001$, $d = 6.93$.

This analysis was replicated with single-subject analyses. For those stimuli that a particular participant judged as typical of Group A, the heights of the intelligence bars were compared with the heights of the friendliness bars. In the B_f condition, 12 of 13 participants showed the predicted pattern (i.e., intelligence $>$ friendliness), $F_s > 5.17$, $ps < .05$. In the B_i condition, 11 of 13 participants showed the predicted pattern (i.e., friendliness $>$ intelligence), $F_s > 6.58$, $ps < .05$. None of the remaining participants showed significant results opposite the predicted pattern.

Discussion

In Study 1, participants' estimates of Group A's central tendency were contrasted from Group B. As the portion of Group A that overlapped with Group B was systematically altered, the participants' assessments of Group A also shifted, even though the Group A stimuli remained constant. This pattern of bias was evident in participants' estimates of the Group A average as well as in the stimuli they judged to be typical of Group A. These biases were demonstrated for participants who distinguished Group A members from Group B members at near optimal levels in the training phase. Thus, as expected, very good categorization performance was accompanied by rather gross errors in judgments about category attributes.

The main effect for trait type on the group average judgments was unanticipated. Averaged over both stimulus types, participants' central tendency estimates had intelligence bars that were

Table 1
Perceived Group A Characteristics by Condition for Different Measures of Central Tendency (Study 1)

Variable	Intelligence		Friendliness		Difference	Cohen's d
	M	SD	M	SD		
Group A average estimates						
B_f	201.2	32.1	122.7	27.1	78.5	2.64
B_i	152.6	48.4	175.3	35.0	-22.7	0.54
Stimuli judged typical of A						
B_f	221.9	8.4	145.2	13.2	76.7	6.93
B_i	155.4	20.1	214.5	17.1	-59.1	3.17

Note. All trait measures are in screen pixels. B_f = B more friendly and less intelligent condition; B_i = B more intelligent and less friendly condition.

higher than the corresponding friendliness bars. Because participants' stereotypes were reliably shifted for both conditions, this main effect does not limit the conclusions of the current study. It does, however, suggest that experiments should be designed so that the crucial model predictions are tested as interactions between traits and other factors.

Note that assessments of a given trait across conditions differ by up to 69 pixels in the predicted direction. Compare this with a true difference across conditions of 0 pixels (i.e., the same Group A stimuli were used in both conditions). In absolute terms, these results indicate that participants' estimates differed by up to 2.4 cm on the computer screen (corresponding to a visual angle for the difference of 2.2°). In addition, the effect sizes were moderate to very large. It is difficult to compare the extent of contrast with a study like Goldstone's (1994), which measured categorical perception using signal detection analyses of same-different judgments. However, the magnitude of the effect in the present study seems larger than one might expect on the basis of perceptual disambiguation processes alone.

Study 2

Study 2 also kept the information about the target Group A constant across conditions but manipulated the degree of overlap rather than the direction of overlap. In both conditions, the Group B stimuli were, on average, more friendly and less intelligent than the Group A members. The B_{far} condition was similar to the B_{f} condition of Study 1. However, the B_{near} condition shifted the Group B stimuli closer to Group A. The portion of the Group A distribution that the optimal responder would misclassify as belonging to Group B was larger in the B_{near} condition than in the B_{far} condition. Therefore, we expected that assessments of Group A's attributes would have a greater bias away from Group B in the B_{near} condition than in the B_{far} condition.

The extent of overlap between Groups A and B also has implications for the perceived variability within a category. Because a larger portion of Group A is misclassified by the optimal responder in the B_{near} condition than in the B_{far} condition, we predicted Group A would be perceived as more homogeneous in the B_{near} condition than in the B_{far} condition. This prediction is all the more interesting because it directly opposes a prediction based on psychophysical arguments. According to Weber's law (e.g., Whitaker, Bradley, Barrett, & McGraw, 2002), as the groups move farther apart, the proportion of the space occupied by Group A gets smaller and Group A should be seen as more homogeneous. Thus, our prediction of more homogeneous beliefs about Group A in the B_{near} condition than in the B_{far} condition would only hold if our predicted effects overpower a well-supported psychophysical law.

Many different measures of perceived homogeneity have been used in past research. However, these measures are often incapable of gauging the perceived distribution of multidimensional stimuli. For instance, a standard method of measuring dispersion is to find out which stimuli are perceived to be the highest and lowest on the desired dimension, thus obtaining the perceived range of the target group. For stimuli such as ours, the range of interest involves neither the most intelligent nor the friendliest members, but rather the instances that simultaneously maximize one dimension while minimizing another (the upper right and lower left portions of the distribution; see Figure 1). Thus, our predictions demanded the use

of an inherently multivariate measure of dispersion. We realized that we could gauge the perceived dispersion by observing the variability in the exemplars that participants judged to be typical of Group A in the exemplar-based measure of central tendency. Each stimulus presented in this task describes a specific combination of intelligence and friendliness, so this measure is inherently multivariate and can get at stimuli that participants see as "less A-like" because they maximize one dimension and minimize another.

Furthermore, because the exemplars presented for typicality judgments were evenly distributed throughout the stimulus space, each represents an equal area (900 pixels²). The number of exemplars categorized as typical of Group A thus provides a multidimensional measure of the perceived dispersion within Group A and is ideal for our study because it integrates the perceived dispersion across both dimensions.

In Study 2, then, the exemplars judged as typical of Group A were used to compute both an exemplar-based measure of central tendency and an exemplar-based measure of dispersion. As before, the bar adjustment measure was also included as an additional test of beliefs about Group A central tendency. We predicted the two measures of central tendency would reveal more contrast away from Group B in the B_{near} condition than in the B_{far} condition. In addition, we expected the exemplar-based measure of dispersion to reveal more within-group homogeneity in the B_{near} condition than in the B_{far} condition.

Method

Design and participants. The design was similar to that of Study 1, except extent of overlap was manipulated rather than direction of overlap. Thus, Study 2 was a 2 (B_{near} vs. B_{far}) \times 2 (bar/exemplar based vs. exemplar based/bar counterbalance) \times 2 (intelligence vs. friendliness) design, with the latter factor being within-subject and the other two factors being between-subjects. Forty undergraduates enrolled in introductory psychology at the University of Illinois, Urbana-Champaign participated in this study. They received partial course credit for their participation.

Stimulus materials. In the B_{far} condition, the Group A and Group B means were the same as those in the B_{f} condition of Study 1 ($A: \mu_{\text{Intelligence}} = \mu_{\text{Friendliness}} = 180$; $B_{\text{far}}: \mu_{\text{Intelligence}} = 125$; $\mu_{\text{Friendliness}} = 235$). To create the B_{near} stimuli, a constant was subtracted from the friendliness scores and added to the intelligence scores of the B_{far} stimuli ($B_{\text{near}}: \mu_{\text{Intelligence}} = 150$, $\mu_{\text{Friendliness}} = 210$). This transformation shifted the B stimuli so that they overlapped more with the A stimuli in the B_{near} condition than in the B_{far} condition, resulting in more stimuli being misclassified by the optimal responder in the B_{near} condition. That is, even though the same Group A stimuli are presented in the two conditions, maximal categorization accuracy is achieved by moving the classification boundary 17.7 pixels closer to the Group A mean in the B_{near} condition as compared with the B_{far} condition (i.e., the y-intercept shifts by 25 pixels). In Study 2, the standard deviation was slightly smaller ($\sigma = 25$) than in Study 1, producing optimal accuracies of 95.7% in the B_{far} condition and 80.2% in the B_{near} condition.¹

¹ The optimal classifier also misclassifies the least friendly and most intelligent Group B members. Assessing beliefs about Group A attributes, however, reveals the effect of the near-far manipulation in the absence of any change in the group's actual attributes. It is possible that the increase in the actual similarity of Group B to Group A in the B_{near} condition would be offset by increased misclassification-driven contrast. We did not test this possibility here.

Procedure. The procedure was the same as that for Study 1 with two exceptions. First, because the two conditions differed in the amount of overlap between categories, the training criteria were altered to reflect the relative difficulty of the two stimulus sets. In the B_{far} condition (5% overlap), participants had to get 91% correct within the previous 150 trials before moving on to the test phase. In the B_{near} condition (20% overlap), participants had to get 76% correct within the previous 150 trials before moving on to the test phase. These criteria were chosen so that participants in both conditions would learn to an accuracy that is within ~4% of an optimal classifier. Note that in both conditions, perceiver accuracy had to exceed the accuracy of the optimal unidimensional classifier (B_{near} optimal unidimensional accuracy = 72.7%; B_{far} optimal unidimensional accuracy = 89.7%). Second, the Study 2 procedure explicitly informed participants that although none of the test stimuli (in the exemplar-based measure) were from Group B, some of them would, nonetheless, be rather atypical of Group A. These instructions emphasized that this task was completely different from the training task.

Results

Replicating Study 1, participants’ beliefs about the central tendency of Group A were expected to be biased in a manner that would increase contrast between the two groups. As in the B_f condition of Study 1 (Figure 1), contrast away from Group B would lead to estimates of intelligence that are higher than estimates of friendliness for Group A. More important, it was predicted that the size of this bias would be greater in the B_{near} condition than in the B_{far} condition because of the higher degree of overlap in the B_{near} condition. In addition, we predicted that the change in category overlap should affect the range of exemplars that participants judged as typical of Group A. Specifically, participants should consider a smaller number of exemplars as typical of Group A in the B_{near} condition than in the B_{far} condition.

Bar height adjustment task. Three participants were excluded from this analysis because they failed to adjust both bars before continuing with the experiment. The Group A average estimates for the remaining 37 participants replicated Study 1 and showed the predicted biases away from Group B (Table 2). Across both experimental conditions, participants’ central tendency estimates had intelligence bars ($M = 174.2$) that were significantly higher than their friendliness bars ($M = 141.9$), $F(1, 33) = 66.4, p < .001, d = 0.99$. A simple-effects analysis was conducted to determine whether the predicted bias occurred reliably within each of the stimulus conditions. The average of the intelligence bars was

reliably higher than the average of the friendliness bars in both the B_{near} and B_{far} conditions, $F(1, 33) = 50.8, p < .001, d = 1.28$ and $F(1, 33) = 15.6, p < .001, d = 0.66$, respectively. Most important, the Extent of Overlap \times Trait interaction was also significant, $F(1, 33) = 6.71, p < .05$. As predicted, the size of the bias increased in the B_{near} condition relative to the B_{far} condition.

Exemplar-based measure of central tendency. Typicality judgments for the 40 participants also revealed the predicted biases (see Table 2). Across both the B_{near} and B_{far} conditions, the exemplars that participants judged typical of Group A had higher average intelligence bar heights ($M = 212.4$ pixels) than friendliness bar heights ($M = 149.7$ pixels), $F(1, 36) = 442.8, p < .001, d = 6.27$. In addition, a simple-effects analysis was conducted to determine whether the predicted bias occurred reliably within each of the overlap conditions. The average of the intelligence bars was higher than the average of the friendliness bars in both the B_{near} and B_{far} conditions, $F(1, 36) = 297.1, p < .001, d = 7.29$, and $F(1, 36) = 145.8, p < .001, d = 5.21$, respectively.

Single-subject analyses also showed the predicted effects. In the B_{near} condition, 20 of 21 participants showed the predicted pattern, $ps < .05$. In the B_{far} condition, 15 of 19 participants showed the predicted pattern, $ps < .05$. One participant did show significant results opposite the predicted pattern, $p < .05$.

The main prediction of Study 2 was confirmed in the exemplar-based measure of central tendency as well, as shown by the significant Extent of Overlap \times Trait interaction, $F(1, 36) = 12.91, p < .05$. As expected, the size of the bias was larger in the condition in which Group B was nearer to Group A.

Exemplar-based measure of dispersion. It was predicted that participants would judge a larger number (i.e., a wider range) of exemplars as typical of Group A when the groups were farther apart. As predicted, participants in the B_{far} condition judged more exemplars as typical of Group A ($M = 50.47$) than did participants in the B_{near} condition ($M = 46.86$), $F(1, 38) = 4.49, p < .05, d = 0.66$. Thus, increased overlap between categories led to a decreased multidimensional range of stimuli that were judged to be typical of Group A.

Stimuli viewed in training. Because the training criteria were different for the two stimulus conditions, it is possible that participants in the two conditions viewed different numbers of training stimuli. There was a marginal trend indicating that participants in the B_{near} condition viewed more stimuli before completing training

Table 2
Perceived Group A Characteristics by Condition for Different Measures of Central Tendency (Study 2)

Variable	Intelligence		Friendliness		Difference	Cohen’s <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Group A average estimates						
B _{far}	175.6	33.2	154.4	30.6	21.2	0.66
B _{near}	172.9	33.0	130.0	33.9	42.9	1.28
Stimuli judged typical of A						
B _{far}	207.6	10.2	155.7	9.7	51.9	5.21
B _{near}	217.2	10.3	143.9	9.8	73.3	7.29

Note. All trait measures are in screen pixels. B_{far} = similar to the B more friendly and less intelligent condition; B_{near} = stimuli shifted closer to Group A.

($M = 184.6$) than did participants in the B_{far} condition ($M = 140.4$), $F(1, 38) = 3.30$, $p < .10$, $d = 0.56$.

Discussion

As in the previous studies, the perceived central tendency of Group A was contrasted away from Group B. More important, the magnitude of this bias was affected by the relative closeness of the two groups such that the bias was largest when the groups were more similar. This difference in bias cannot be easily explained as a function of the different learning criteria used in the two conditions because the criteria for the B_{near} condition required participants to view slightly more training stimuli. If anything, this increased training should have made the estimates in the B_{near} condition more accurate than those in the B_{far} condition, opposite the obtained pattern of results.

Also as predicted, perceivers' beliefs about Group A attributes were more homogeneous as the true intergroup similarity increased and more Group A stimuli were optimally misclassified as belonging to Group B. If anything, this finding is counterindicated by well-established laws of psychophysics. Psychophysical effects drive perceivers to see the same two stimuli as less similar to one another when the total range of the stimuli is small and more similar to one another when the total range of stimuli is large. Perceptual effects would thus drive perceivers to see the Group A members as less homogeneous in the B_{near} condition. Study 2 produces the opposite effect. Thus, categorization effects are presumably overpowering any psychophysical scaling effects in this study.

The findings of Study 2 have both positive and negative implications for accuracy of knowledge about category attributes. On the plus side, the biases are not pervasive because they have minimal effects for highly distinct groups. On the minus side, the biases that do exist are resistant to elimination because a reduction in actual group differences is partially compensated by increased contrast. In short, perceived central tendency and variability are warped in a way that maintains group distinctions. The resulting biases may not always be large, but are likely to be persistent.

Study 2 demonstrated that both the between-categories contrast and the within-category assimilation are dependent on the objective extent of overlap in the distributions of the groups. The more similar the groups are, the larger both biases become. This is consistent with the idea that biases in beliefs about category attributes increase as the portion of the category distribution that is misclassified by the optimal observer increases. In summary, taking the perspective of the optimal classifier suggests an interplay between perceived central tendency and perceived variability that is substantiated by the experimental data.

Study 3

In Study 3, the misclassifications made by the optimal responder were manipulated by changing the relative base rates for the two groups. Thus, Group A was either a majority compared with Group B or a minority compared with Group B. Although the relative sizes of the groups varied, the means and variances of the groups remained constant across conditions.

When base rates are equal, classification accuracy is maximized by a classification criterion that is defined by the points at which

stimuli are equally likely to belong to either category. However, if the number of members in one category is increased, stimuli that had a 50/50 chance of belonging to either category are now more likely to belong to the more numerous category. Therefore, the optimal decision criterion shifts to include these stimuli in the more numerous category (Ashby, 1992; Ashby & Gott, 1988; Maddox, 1995). Although this shift produces more correct classifications overall, it increases the misclassifications of stimuli in the overlapping portion of the minority category. This should result in a larger contrast effect for a minority group than for a majority group.

In addition, factors that move the optimal classification boundary toward a category's mean narrow the range of stimuli that people classify as belonging to that category. As mentioned above, when the base rates of two categories are made unequal, the decision criterion shifts toward the less numerous category. Therefore, perceivers should see a minority group as more homogeneous than a majority group.

Method

Design and participants. The design of Study 3 was a 2 (1:2 vs. 2:1 base rate) \times 2 (bar/exemplar based vs. exemplar based/bar counterbalance) \times 2 (intelligence vs. friendliness) factorial, with the latter factor manipulated within subject and the other two factors manipulated between subjects. Twenty-seven undergraduates enrolled in introductory psychology at the University of California, Santa Barbara participated in this study. They received partial course credit for their participation.

Stimulus materials. The values of the training stimuli in this experiment were taken from the B_f condition of Study 1 (10% overlap) and were presented in the same manner. Unlike Study 1, however, the number of stimuli shown from the two stimulus groups was not equal. This was accomplished by creating two sets of 750 training stimuli. In one set, each of the 250 Group A stimuli was included twice, whereas the 250 B_f stimuli were included only once (2:1 condition). In the other set, each of the Group A stimuli was included once, whereas the B_f stimuli were included twice (1:2 condition). Even though the stimuli for each group have the same means, variances, and covariance across the base-rate conditions, the optimal bound shifts ~ 16 pixels toward the Group A mean in the 1:2 condition as compared with the 2:1 condition. (Put a different way, the y -intercept is ~ 22 pixels higher in the 2:1 condition than in the 1:2 condition.)

Procedure. The procedure of the current study was identical to that of Study 2 with the exception that the two conditions differed in the base rates between groups rather than in the means of Group B stimuli. In addition, consistent with the use of the B_f stimuli from Study 1, the Study 1 learning criterion was used in the present study (87.3%). This learning criterion was below optimal accuracy (90%) but exceeded the maximum accuracy that could be attained using a unidimensional rule (84.8%).

Results

With these stimuli, the central tendency predictions would be confirmed if perceivers' judgments about Group A's intelligence were higher than their judgments about friendliness and if this effect were particularly pronounced in the 1:2 base rate condition. The variability predictions would be confirmed if Group A was seen as more homogeneous in the 1:2 base rate condition than in the 2:1 base rate condition.

Bar height adjustment task. Four participants were excluded from this analysis because they failed to adjust both bars. The group average judgments for the remaining 23 participants showed

the predicted biases (Table 3). Across both base rate conditions, participants' central tendency estimates had intelligence bars ($M = 177.8$) that were significantly higher than their friendliness bars ($M = 146.2$), $F(1, 19) = 65.8$, $p < .001$, $d = 1.29$. A simple-effects analysis was conducted to determine whether the predicted bias occurred reliably within each of the base rate conditions. The average of the intelligence bars was reliably higher than the average of the friendliness bars in both the 1:2 and 2:1 conditions, $F(1, 19) = 35.1$, $p < .001$, $d = 1.48$, and $F(1, 19) = 30.2$, $p < .001$, $d = 1.01$, respectively.

There was a trend in the predicted direction for the Trait \times Base Rate interaction, $F(1, 19) = 3.516$, $p < .10$. Thus, there is tentative evidence that the size of the bias is increased in the 1:2 base rate condition relative to the 2:1 condition.

Exemplar-based measure of central tendency. Three participants were excluded from this analysis: 2 because they held down a single response key for a significant number of consecutive trials, and 1 because he failed to follow the task instructions. Typicality judgments for the remaining 24 participants showed the predicted biases (Table 3). Across both base rate conditions, the exemplars that participants judged typical of Group A had higher average intelligence bar heights ($M = 224.6$ pixels) than friendliness bar heights ($M = 139.5$ pixels), $F(1, 20) = 1343.00$, $p < .001$, $d = 14.47$. In addition, a simple-effects analysis was conducted to determine if the predicted bias occurred reliably within each of the base rate conditions. The average of the intelligence bars was higher than the average of the friendliness bars in both the 1:2 and 2:1 conditions, $F(1, 20) = 33.78$, $p < .001$, $d = 13.35$, and $F(1, 20) = 30.4$, $p < .001$, $d = 15.42$, respectively.

This analysis was replicated by comparing mean intelligence and friendliness bar heights in single-subject analyses. In the 1:2 condition, 15 of 15 participants showed the predicted pattern, $ps < .05$. In the 2:1 condition, 8 of 9 participants showed the predicted pattern, $ps < .05$. The remaining participant did not produce results significantly opposing the predicted effects.

The main prediction of Study 2 was confirmed, and the Base Rate \times Trait interaction was also significant, $F(1, 20) = 5.57$, $p < .05$. The size of the bias was larger in the 1:2 base rate condition than in the 2:1 condition.

Exemplar-based measure of dispersion. It was predicted that participants would judge a larger number (i.e., a wider range) of exemplars as typical of Group A when Group A was in the majority than they would when Group A was in the minority. As

predicted, participants in the 2:1 condition judged more exemplars as typical of Group A ($M = 76.83$) than did participants in the 1:2 condition ($M = 67.54$), $F(1, 20) = 13.14$, $p < .01$, $d = 1.34$. These data are consistent with the prediction that minority status leads to increased perceptions of homogeneity because minority status is accompanied by optimal misclassification of a larger portion of the category. However, these results can also be explained via one of at least three other mechanisms. First, a perceiver who saw more stimuli in a group might feel compelled to report more members as being typical of the group. With our uniformly distributed test stimuli, this would produce a wider measure of variability for the majority compared with the minority group. Second, a perceiver might use a straight measure of numerosity to determine typicality (e.g., "If I saw more than 3 of them, it's typical"). Third, a perceiver's judgments of homogeneity might be affected by group size because even if two groups have identical variability in reality, the smaller group provides relatively less exposure to members at the extremes of the category distribution (Linville, Fischer, & Salovey, 1989). It is true that any of these alternative explanations would produce enhanced homogeneity for the minority group relative to the majority group in the present study. It is also true that none of these alternative explanations rely on misclassification of the optimally misclassified stimuli. However, these explanations can account neither for the central tendency results nor for the dispersion results of Study 2. Only the misclassification argument can parsimoniously explain the combined results of Studies 1 through 3.

Discussion

In Study 3, higher intelligence than friendliness ratings indicated that perceivers contrasted beliefs about Group A's attributes away from those of Group B. More important, the size of the bias was predicted to be larger in the condition in which Group A was less numerous than Group B. This effect was found using the exemplar-based measure of central tendency. A trend in this direction was also found in participants' bar height adjustment estimates of the Group A average. Finally, participants judged a narrower range of exemplars as being typical of Group A when members of that group were in the minority in the training phase of the experiment. Overall, this pattern of effects is consistent with predictions: Increasing the number of optimally misclassified

Table 3
Perceived Group A Characteristics by Condition for Different Measures of Central Tendency (Study 3)

Variable	Intelligence		Friendliness		Difference	Cohen's d
	M	SD	M	SD		
Group A average estimates						
2:1 A:B	167.9	24.1	144.4	22.3	23.5	1.01
1:2 A:B	186.3	25.5	148.0	26.2	38.3	1.48
Stimuli judged typical of A						
2:1 A:B	220.8	5.9	141.2	4.3	79.6	15.42
1:2 A:B	228.4	5.3	137.8	8.0	90.6	13.35

Note. All trait measures are in screen pixels.

Group A stimuli led to increases in both between-categories contrast and within-category assimilation.

The results of Study 3 also have practical implications for understanding the perception of minority and majority groups. In spite of the fact that participants were trained to a high accuracy criterion in the categorization task, their assessments of the minority group demonstrated greater bias and greater perceived homogeneity relative to the majority group. This effect occurred in the absence of any explicit group affiliation for the participants, and it cannot be explained as a general positivity bias for majority group members because the bias was positive on one dimension and negative on the other. This evidence for a nonmotivated bias in one's perception of minority groups has implications for a wide range of social problems, for example, unequal allocation of resources and the maintenance of intergroup conflict. It also suggests that even the most well-informed and well-intended people may find it difficult to accurately perceive minority groups.

Study 4

Studies 1 through 3 are consistent with the idea that altering the comparison category so that a larger portion of the target category is optimally misclassified increases between-categories contrast and within-category assimilation. To this point, however, any other mechanism that involves similarity between two categories might also explain our results. The goal of Study 4 was to strengthen the argument that optimized misclassification is driving our effects. Study 4 investigated assessments of a fixed-target group in two diagnostic situations. Both situations involved the same degree of intercategory overlap, but one situation included a goal to maximize classification accuracy, whereas the other situation did not.

All participants in Study 4 were presented with identical information about the scholastic abilities of members of four groups, Groups A, B, C, and D (Figure 2). In the ABCD condition, participants decided to which of the four groups each stimulus belonged with the benefit of corrective feedback on each trial. In the other two conditions, they classified Groups A and B with feedback and, separately, classified Groups C and D with feedback. In one of these two latter conditions, participants learned about Groups A and B first (the AB_CD condition). In the other, participants learned about Groups C and D first (the CD_AB condition). Thus, the goal of optimizing Group B versus Group C classification accuracy is present in the ABCD condition but is absent in the AB_CD and CD_AB conditions.

After classifying all of the stimuli, each participant assessed the attributes of Group B. An optimal classifier in the ABCD condition would misclassify both the lowest ability Group B members (as

Group A members) and the highest ability Group B members (as Group C members). In contrast, the optimal classifier in the AB_CD and CD_AB conditions would misclassify only the lowest ability Group B members (as Group A members). If optimized misclassification contributes to contrast and assimilation, then the ABCD condition should produce lower estimates of perceived central tendency compared with the AB_CD and CD_AB conditions. In addition, Group B should be perceived as more homogeneous in the ABCD condition than in the AB_CD and CD_AB conditions. This difference in perceived homogeneity should be driven by the high end of the Group B distribution.

An alternate possibility is that overlap with similar categories impacts assessments of a group's attributes regardless of whether systematic misclassification of stimuli has occurred. If so, assessments of Group B's attributes would not vary as a function of learning condition.

Method

Design and participants. The design of Study 4 is a 3 (ABCD vs. AB_CD vs. CD_AB learning condition) \times 2 (bar/exemplar based vs. exemplar based/bar counterbalance) \times 2 (low/high bar vs. high/low bar counterbalance) between-subjects factorial. The learning condition factor manipulated the variable of interest, whereas the other two factors served only as counterbalances. Two hundred thirty-nine undergraduates enrolled in introductory psychology at Indiana University participated in this study. They received partial course credit for their participation.

Stimulus materials. Each stimulus consisted of a single vertical white bar displayed against a black background. The height of the bar indicated the scholastic ability score for a single individual. For the training phase of the experiment, 90 stimuli were sampled from each of four univariate normal distributions (Figure 2). The means of the four groups were ordered in 70-pixel increments: $\mu_A = 115$ pixels; $\mu_B = 185$ pixels; $\mu_C = 255$ pixels; and $\mu_D = 325$ pixels. For all groups, the standard deviation was equal to 30 pixels. Random samples of 90 stimuli were drawn from each of these four populations. The sampled stimuli for each group were then mathematically transformed to match the population mean and population standard deviation. The amount of overlap per tail was approximately 9.1%.

The same set of 90 Group A, 90 Group B, 90 Group C, and 90 Group D training stimuli was shown to participants in all conditions. The order of presentation of the stimuli within the training phase was randomly determined. The stimuli were presented to all participants in the ABCD condition in the same random order. The Group A versus Group B training set was generated by removing the Group C and Group D stimuli from this ordered ABCD training set. The Group C versus Group D training set was generated by removing the Group A and Group B stimuli from this ordered ABCD stimulus set.

For the test phase of the experiment, 88 stimuli were sampled without replacement from a uniform distribution ranging from a bar height of 56 to

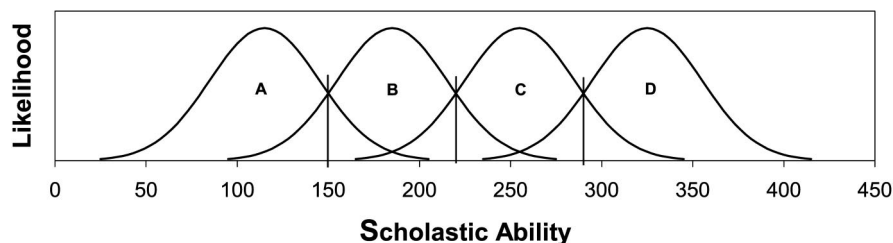


Figure 2. Univariate normal stimulus distributions for Groups A, B, C, and D of Study 4.

a bar height of 400. All participants judged the typicality of each test stimulus. The order of presentation of the test stimuli was randomized but held constant for all participants.

Procedure. Participants learned that they would view scholastic ability scores for a series of individuals and that their task would be to select the group to which each individual belonged. In the ABCD condition, participants were instructed that they would see members of four different groups and that they should press one of four response keys labeled A, B, C, or D to indicate the group membership corresponding to the presented scholastic ability score. After each classification choice, participants received corrective feedback and continued on to the next trial.

In the AB_CD and CD_AB conditions, participants received similar instructions but were told they would learn about two groups. In the AB_CD condition they first learned about Groups A and B, and in the CD_AB condition they first learned about Groups C and D. They were presented with the corresponding stimuli. Participants pressed one of two labeled keys to indicate their classification choice for each stimulus. Feedback was provided after each classification attempt before the participant continued on to the next trial. After completing training on two groups, participants in the AB_CD and CD_AB conditions repeated the same procedure with the two groups they had not yet classified.

All participants trained once on each of the 360 training stimuli and then proceeded to the test phase. In the test phase, participants completed an exemplar-based measure and a group average bar adjustment task similar to those in Studies 1 through 3 (except they included only the one dimension of scholastic ability). The order of the exemplar-based measure and the bar height adjustment task was counterbalanced. In Study 4, participants also adjusted bar heights to indicate their estimates of the lowest Group B member and of the highest Group B member. The low and high estimates always followed the group average estimate, and the order of the low and high estimates was counterbalanced.

Results

If misclassification plays a role in assessments of group attributes, assessments of Group B's central tendency and dispersion should be lower in the ABCD condition than in the AB_CD and the CD_AB conditions. Results are presented in Table 4. As in previous studies, the counterbalancing produced no significant effects.

Bar height adjustment task. Group average judgments for Group B showed the predicted biases. The ANOVA comparing all three conditions was significant, $F(2, 236) = 3.93, p < .021$. A contrast indicated that the estimate of group average in the ABCD condition ($n = 72, M = 135.11, SD = 57.36$) was lower than the

estimate of group average in the AB_CD and CD_AB conditions ($n = 167, M = 160.51, SD = 69.22$), $F(1, 236) = 7.59, p < .006, d = 0.384$. A contrast comparing the AB_CD and CD_AB conditions failed to reach significance, $p > .5$

High and low assessments in the bar height adjustment task. Participants also adjusted a bar to the highest Group B scholastic ability. As expected, these estimates also significantly differed as a function of condition, $F(2, 236) = 19.63, p < .0001$. A contrast indicated that the high estimate in the ABCD condition ($n = 72, M = 179.00, SD = 43.01$) was lower than the high estimate in the AB_CD and CD_AB conditions ($n = 167, M = 249.02, SD = 92.73$), $F(1, 236) = 36.86, p < .0001, d = 0.861$. The AB_CD and CD_AB conditions did not significantly differ, $p > .19$.

Unexpectedly, participants' estimates of the lowest Group B scholastic ability also significantly differed by condition, $F(2, 236) = 3.73, p < .03$. A contrast indicated that the low estimate in the ABCD condition ($n = 72, M = 94.78, SD = 27.66$) was lower than the low estimate in the AB_CD and CD_AB conditions ($n = 167, M = 112.11, SD = 50.61$), $F(1, 236) = 7.46, p < .007, d = 0.384$. The low estimates in the AB_CD and CD_AB conditions did not significantly differ, $p > .87$.

Exemplar-based measure of central tendency. One participant was excluded from the analyses for judging all of the exemplars as typical of Group B. Typicality judgments for the remaining 238 participants showed that the learning conditions significantly affected the average of the scholastic abilities judged typical of Group B, $F(2, 235) = 25.49, p < .0001$. A contrast indicated that the mean of the abilities rated as typical of Group B in the ABCD condition ($n = 72, M = 173.04, SD = 26.05$) was lower than the corresponding mean of the AB_CD and CD_AB conditions ($n = 166, M = 229.11, SD = 64.18$), $F(1, 235) = 50.89, p < .0001, d = 1.006$. The mean in the AB_CD and CD_AB conditions did not significantly differ, $p > .97$.

Exemplar-based measure of dispersion. Because the exemplars in the typicality task were randomly sampled from a uniform distribution, the stimuli in Study 4 were not evenly distributed across the stimulus space. Thus, the number of exemplars classified as typical of Group B was not a direct measure of dispersion as it was in Studies 1 through 3. Instead, the standard deviation of the scholastic abilities that each participant judged as typical of Group B was analyzed. The standard deviation of those abilities

Table 4
Perceived Group B Characteristics by Condition for Different Measures of Central Tendency (Study 4)

Variable	ABCD		AB_CD		CD_AB		Contrast <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Group average estimates	135.11	57.36	164.03	71.52	157.51	67.44	0.38
Group high estimates	179.00	43.01	240.29	93.32	256.49	92.02	0.86
Group low estimates	94.78	27.65	112.70	56.37	111.60	45.42	0.38
<i>M</i> judged typical	173.04	26.05	228.99	68.58	229.21	60.50	1.01
<i>SD</i> judged typical	41.76	14.16	59.64	16.88	62.02	17.27	1.17
Maximum judged typical	267.11	62.98	350.05	63.16	351.33	65.99	1.30
Minimum judged typical	96.64	29.70	114.26	60.54	105.53	43.41	0.28

Note. All trait measures are in screen pixels. Contrast *d* is the effect size for the ABCD versus AB_CD and CD_AB contrast, all significant at $p < .01$.

rated as typical of Group B was significantly affected by the learning condition, $F(2, 235) = 35.28, p < .0001$. A contrast indicated that the standard deviation of the abilities rated as typical of Group B in the ABCD condition ($n = 72, M = 41.76, SD = 14.16$) was lower than the corresponding standard deviation for the AB_CD and CD_AB conditions ($n = 166, M = 60.91, SD = 14.16$), $F(1, 235) = 68.94, p < .0001, d = 1.174$. The standard deviation of the abilities rated as typical in the AB_CD and CD_AB conditions did not significantly differ, $p > .34$.

High and low assessments in the exemplar-based measure. The high and low assessments were also analyzed because the expectation was that assessments of the high end of the Group B distributions would be affected by the learning condition but that assessments of the low end would not. Partially confirming predictions, the highest ability rated as typical of Group B varied with learning condition, $F(2, 235) = 42.63, p < .0001$. A contrast indicated that the highest ability rated as typical of Group B in the ABCD condition ($n = 72, M = 267.11, SD = 62.97$) was lower than the highest ability rated as typical in the AB_CD and CD_AB conditions ($n = 166, M = 350.73, SD = 62.97$), $F(1, 235) = 85.02, p < .0001, d = 1.302$. The highest abilities rated as typical in the AB_CD and CD_AB conditions did not significantly differ, $p > .89$.

The lowest ability rated as typical of Group B was marginally affected by the learning condition, $F(2, 235) = 2.68, p < .07$. Unexpectedly, however, a contrast indicated that the lowest ability rated as typical of Group B in the ABCD condition ($n = 72, M = 96.64, SD = 29.70$) was lower than the lowest ability rated as typical in the AB_CD and CD_AB conditions ($n = 166, M = 109.58, SD = 52.08$), $F(1, 235) = 4.08, p < .04, d = .277$. The lowest abilities rated as typical in the AB_CD and CD_AB conditions did not significantly differ, $p > .22$.

Discussion

In Study 4, not only did participants learn about the same stimuli from the target group (Group B), but they also learned about the same stimuli from comparison Groups A, C, and D. Despite this equivalence of presented stimuli, participants' assessments of Group B attributes differed greatly depending on whether they categorized Group B in direct comparison with Group C or not. An optimal responder comparing Groups B and C would misclassify the most able Group B members as belonging to Group C. Consequently, optimizing categorization of Groups B versus C (ABCD condition) was expected to lead to assessments of Group B as relatively less scholastically able as compared with conditions in which no such misclassification of the most able Group B members occurred (AB_CD and CD_AB).

As predicted, when Group B was directly categorized in comparison to Group C, Group B was seen as both less scholastically able and less variable in scholastic ability. Also as expected, the difference in variability was largely due to differences in assessments of the most able Group B member (Cohen's $d = 1.30$) rather than to differences in assessments of the least able Group B member (Cohen's $d = 0.28$).

Study 4 clearly shows that it is not just the relative likelihood that affects assessments of category attributes, but also the relative likelihood of those categories with which the target category is compared when optimizing classification accuracy. This is consis-

tent with the view that misclassification during categorization learning contributes to between-categories contrast and within-category assimilation.

General Discussion

The claim of the present research is that optimizing classification accuracy can produce systematic misclassifications that contribute to both between-categories contrast and within-category assimilation. Consistent with this hypothesis, four manipulations that altered the behavior of an optimal classifier produced predicted shifts in perceivers' assessments of category attributes. In Study 1, the direction of overlap with the comparison category systematically altered whether (a) the most friendly and least intelligent members or (b) the least friendly and most intelligent members of the target category were optimally misclassified. As predicted, this manipulation altered perceivers' assessments of the target category's attributes in a manner that underutilized optimally misclassified stimuli and exaggerated between-groups differences. In Study 2, increased overlap with the comparison category increased the portion of the target category that was optimally misclassified. The result was a corresponding increase in both between-categories contrast and within-category assimilation. Study 3 demonstrated the predicted effect of relative base rate. Optimal categorization is achieved by shifting the categorization boundary closer to the minority group mean and further from the majority group mean, producing increased misclassification of those minority group members that are most similar to the majority group. As predicted, this shift was accompanied by an increased bias in beliefs about minority group central tendency and increased assessments of minority group homogeneity. In Study 4, between-categories overlap that involved optimized misclassification produced different assessments of group attributes than did overlap that did not involve optimized misclassification. As expected, the direction of this effect supported the contention that optimally misclassified stimuli have attenuated impact on assessments of group attributes.

It is interesting that Studies 2 through 4 consistently and simultaneously produced both between-categories contrast and within-category assimilation on categorization-relevant dimensions with effect sizes that were moderate to large. These findings stand in contrast to prior work that has suggested the simultaneous occurrence of between-categories contrast and within-category assimilation is rare. Unlike the few studies that have reported simultaneous contrast and assimilation (Goldstone et al., 2001; Livingston & Andrews, 2005), our studies show both effects using a single measure. Presumably, obtaining both effects on a single measure provides somewhat stronger evidence that a single mechanism contributes to both effects than would be the case if each effect was obtained with a different measure. In addition, the present results clearly confirm that categorization can, indeed, bias assessments of category attributes. In fact, highly accurate categorization learning can lead to very misguided views about what a category is like.

Comparisons to Prior Work

There are three notable differences between the present work and prior work on between-categories contrast and within-category

assimilation. First, in terms of theory, we suggest that systematic misclassification that boosts overall categorization accuracy can lead to predictable distortions in beliefs about what a category is like. Various mechanisms could account for why this might occur. Some possibilities include (a) storing optimally misclassified exemplars with the wrong category label, (b) storing a truncated categorization response region as the representation of the category, (c) storing only the decision bound that optimizes categorization accuracy, (d) weakening (or strengthening) connections between the optimally misclassified stimuli and the correct (or incorrect) category label, or even (e) weighting the optimally misclassified stimuli less at the time of judgment. Clearly, it is beyond the scope of this article to differentiate between these representational assumptions. The key point here is that a variety of theoretical instantiations could all lead to the same previously uninvestigated claim that, ironically, accurate categorization can lead to an outcome of strong and systematic biases in beliefs about categories.

Second, because the premise was that biased beliefs ensue to the extent that systematic misclassification occurs, the category structures in the present studies differed markedly from those used in much of the prior work on between-categories contrast and within-category assimilation. Specifically, the current studies used large, normally distributed categories that overlapped on continuously valued attribute dimensions. Although this is a decided departure from prior work (Corneille & Judd, 1999; Corneille et al., 2002; Eiser, 1971; Goldstone, 1994, 1996; Krueger & Clement, 1994; Krueger et al., 1989; Livingston et al., 1998; McGarty & Penny, 1988; Medin, Wattenmaker, & Hampson, 1987; Rosch & Mervis, 1975; Tajfel & Wilkes, 1963), it is a realistic assumption about what categories are often like in the real world.

Third, in contrast to much social psychological work, motivated biases in group perception are minimized in the present work. This was done by choosing groups to which the perceiver did not belong and about which the perceiver had no preexisting expectations. Consequently, the perceiver could not satisfy competing assimilative and distinctiveness needs (Brewer, Dull, & Lui, 1981), could not satisfy self-enhancing needs (Mullen et al., 1992), and had no need to be consistent with prior expectations (Krueger et al., 1989).

Obviously, other researchers have demonstrated between-categories contrast and within-category assimilation in the absence of any overlap between categories. A variety of mechanisms have been proposed in the literature that might account for contrast and assimilation in these cases. As mentioned in the previous paragraph, motivated responding might produce between-categories contrast and within-category assimilation, even in the absence of intercategory overlap.

Perceptual effects also contribute to between-categories contrast and within-category assimilation, even in the absence of category overlap (Corneille & Judd, 1999; Goldstone, 1994, 1995; Livingston et al., 1998.) There seem to be three main explanations for how perception might be altered as a function of category membership. One is that stimuli that have a common category label seem more similar to one another than the same stimuli with different category labels or with no category labels (Tversky, 1977). This type of perceptual effect explains within-category assimilation. Another argument is that increased attention to a categorization-relevant dimension produces "expansion" along

that dimension so that stimuli seem relatively more different from one another along that dimension (Goldstone, 1994; Nosofsky, 1986). This could lead to intercategory contrast, but it should also lead to the opposite of within-category assimilation. A third mechanism is expansion of the categorization-relevant dimension that is localized at the intercategory boundary. By this account, stimuli near the boundary are easier to discriminate from close neighbors than are stimuli far from the boundary. Note that the label effect and a global expansion of the relevant dimension compete with regard to within-category homogeneity. Although attention to the categorization-relevant dimension produces enhanced discriminability across the dimension, the effect of a common label decreases discriminability. Together, these two effects could produce a result that looks like localized expansion of a dimension at the intercategory boundary (as in Goldstone, 1994.) These competing mechanisms might explain some of the difficulty in consistently obtaining both between-categories contrast and within-category assimilation. Variables that might affect the trade-off between these mechanisms include the strength of similarity conveyed by the category labels (A or B might convey less similarity than extrovert or introvert) and the difficulty of the categorization (more attention to the categorization-relevant dimension is required for more difficult categorization tasks).

Our misclassification explanation for why intercategory contrast and within-category assimilation occur does not negate the importance of these prior contributions but, instead, adds to them. The misclassification argument (a) suggests an additional mechanism that could lead to assimilation and contrast, (b) seems reasonable given the prevalence of overlapping categories in the world and the relative dearth of overlapping categories in the literature on contrast and assimilation, and (c) can produce very large and coincident between-categories contrast and within-category assimilation effects. In fact, the effects of optimized misclassification are potentially very large and may, in many instances, override any perceptually driven effects.

Turning to the realm of social psychology, an interesting comparison can be made between our work and the inclusion/exclusion work of Bless and Schwarz (1998). They argued that assessments of a social group can be altered by manipulating whether a member of that group is temporarily classed as belonging to the group or not. In one condition, perceivers rated the Social Democrat party after thinking about the president in his party-unaffiliated status as president. In the other condition, perceivers rated the Social Democrat party after thinking about the president in his prior role as a long-standing member of the Social Democrat party. They found that perceivers who thought about the target in his party-unaffiliated, presidential role did not include the president's attributes in their assessment of the party's attributes. Thus, classifying the target as president rather than party member led to the exclusion of the target's attributes when judging what the party was like. This sounds quite similar to our proposal that misclassified members' attributes underinfluence beliefs about their group. However, two differences between these approaches are noteworthy. First, our optimal misclassification claims are more quantitatively specified and are integrated within established classification learning and modeling frameworks. Second, and more important, our optimal misclassification claim is driven by biased encoding rather than by context at the time of judgment. If, for example, calling a Group A member a "B" leads to a weakened

association in memory with the “A” label and a strengthened association with the “B” label, these associations will persist beyond the context of a particular task. Thus, optimal misclassification may create biases that are relatively stable and long lasting.

As mentioned earlier, additional learning will not alleviate these biases if the relative likelihoods remain unchanged. The contribution of the systematic misclassification mechanism to biased assessments of category attributes will, however, fluctuate as a function of any factor that alters the optimal classification boundary. Study 4 demonstrated that its contributions also fluctuated as a function of the observer’s motivation to accurately classify the stimuli.

Implications for Stereotyping Research

Certainly the results of these studies have implications for how biased stereotypes might develop. Even in the absence of any motivation to distance beliefs about the attributes of one’s own group from those of another group, biased beliefs can result as a simple function of category learning (Allport, 1954; Krueger et al., 1989). Our knowledge about how this might happen is expanded by the present research. If the attributes of two groups overlap and perceivers rely on that attribute information to assign individuals to groups, then perceivers will think the groups are more different than they really are and that each group is more homogeneous than it really is. More alarming, the present research suggests that accurate category learning might lead to perceptions of numerical minority groups that are highly biased. Any small differences between groups will be exaggerated, and this will largely come at the expense of accuracy in perceptions of the minority group. As mentioned earlier, these biases are likely quite difficult to eliminate: Additional learning does not change the relative likelihoods of the two groups, and consequently, bias will persist. It is important to note that the overlapping characteristics do not have to be evaluative—they could be a simple physical trait such as nose length or lip size. Even if the characteristics are valenced, a minority group could be misperceived as either better than they actually are (e.g., Asians and math) or worse than they actually are (e.g., African Americans and scholastic ability).

Speculatively, one could also consider threat as a negative payoff and make predictions about biased stereotypes within the present framework. If one group is actually more threatening than another, the cost of misclassifying a threatening group member as belonging to the unthreatening group is higher than the cost of misclassifying an unthreatening group member as belonging to the threatening group. This discrepancy in negative payoffs should shift the classification boundary toward the mean of the less threatening group (Maddox & Bohil, 2003). The result would be relatively accurate beliefs about the threatening group but quite biased beliefs about the less threatening group. To use a very loaded example, if African Americans actually were more violent than European Americans, perceptions of African Americans’ levels of violence might be relatively accurate, but beliefs about European Americans’ levels of violence would be underestimated. Possibly more interesting is the idea that other attributes that differ between threatening and unthreatening groups might also be biased by the shift in the classification boundary that occurs as a function of threat. For example, if African Americans actually

were more violent than European Americans, beliefs about African American skin tone might be relatively accurate, but beliefs about European American skin tone might be lily white.

The extent to which optimizing classification accuracy influences beliefs about social groups in the real world remains an open question. It might be argued that one does not learn social categories by classifying a stimulus and then receiving corrective feedback. Although people often do not have immediate trial-by-trial classification feedback, we do manage to develop expectations about the relative likelihoods of different groups, given certain attribute dimensions. Suppose I know, for example, that men who are as effeminate as my friend, Ray, are usually gay. If I encounter a new man who is as effeminate as Ray, I will be more likely to be accurate if I class him as gay than if I class him as straight. I am not likely to ask him outright about his sexual orientation, so I will not get feedback. Instead, I will continue to think of him as gay, regardless of whether this is an accurate or an inaccurate classification. If my classification happens to be correct, I will reaffirm any differences between gay and straight men on the dimension of effeminacy. If it happens to be wrong, I will erroneously add to my repository of beliefs that gay men are effeminate and, simultaneously, erroneously add to my beliefs that straight men are not effeminate. Through this type of optimization of classification on the basis of prior beliefs about likelihoods, I will develop biased impressions that gays are more effeminate than they really are and that straights are less effeminate than they really are. This resonates with prior work on expectancy-based illusory correlation (Hamilton & Rose, 1980) but adds a framework of categorization involving optimized misclassification that reinforces those expectancies. As mentioned earlier, the misclassification framework provides hints about what representational or decisional processes might be involved in generating this bias.

The present studies suggest that attempts to maximize classification accuracy for large, overlapping groups can lead to between-categories contrast and within-category assimilation. We argue that these effects are not driven by altered abilities to perceptually discriminate (e.g., compression and expansion effects) but, rather, are a byproduct of misclassification when attempting to maximize classification accuracy. Even presuming perceivers have accurate beliefs about the likelihoods of different groups, biased beliefs about attributes should persist to the extent that the perceivers engage in optimally classifying individuals to different groups. These biases will only be exacerbated in the absence of corrective feedback following an incorrect classification.

Categorization and Representation

Categorization training involves learning to maximize classification accuracy by assigning stimuli to the correct categories as often as possible. Ashby and colleagues (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Casale, 2003) have suggested that categorization can be optimized in some cases by ascertaining the explicit, verbalizable rule that best separates the categories (see also Nosofsky & Palmeri, 1998). They suggested that such rules are learned via the anterior cingulate gyrus and prefrontal cortex and that people can describe these rules as the basis of their categorization behavior. Based on the hypothesized brain areas involved, they proposed and demonstrated that categorization tasks that can be optimized through explicit rule learning are compro-

mised by including a coincident task that taps the prefrontal cortex, are not affected by the amount of feedback delay, do not suffer when response keys are switched, and can continue with some success in the absence of feedback.

Obviously, such verbal rules can produce optimal categorization for those categories that can be best distinguished by setting a criterion along a single dimension. It is important to note that only knowledge of this optimal rule is required to maximize classification accuracy. If the rule was all that was stored about the categories, perceivers would certainly ignore the optimally misclassified portion of each category. However, such a rule-based representation would provide perceivers with little basis for estimating the group central tendency and dispersion. Perceivers in Study 4's AB_CD and CD_AB conditions seemed to have exactly this kind of inaccuracy in their typicality ratings. In fact, these perceivers rated stimuli as being typical of Group B that were up to 5.5 standard deviations away from the Group B mean!

In Studies 1 through 3, in contrast, optimal classification was achieved if the information from both attribute dimensions was integrated in a manner that could not be easily verbalized. Ashby and colleagues (Ashby et al., 1998; Ashby & Casale, 2003) claimed that such information-integration category structures lead to implicit classification learning in which perceivers learn via a reward-mediated pathway in the striatum (and, more specifically for visual stimuli, in the tail of the caudate nucleus). This type of learning does not allow perceivers to report the basis of their categorization judgments, is not interfered with by a coincident task involving prefrontal cortex, is degraded by feedback delay, suffers if the response keys are switched after training, and is horribly degraded by lack of supervision during training.

Information-integration category learning might be represented in exemplar form (Nosofsky, 1988). On the basis of neural arguments, however, Ashby and colleagues (e.g., Ashby et al., 1998) have suggested two representational possibilities that could account for implicit learning of information-integration category structures. One possibility is that the decision rule itself is represented. A striatal network determines on which side of the decision bound a particular stimulus falls and suggests the appropriate categorical response. Another possibility is that repeated exposure produces learning that maps category labels to sets of neurons that correspond to regions of the perceptual space. In this case, each category label becomes associated with different subregions of the perceptual space.

These possibilities have different implications for how perceivers assess category attributes. First, although exemplars might account for underutilization of stimuli in overlapping portions of the category distribution that are similar to the contrast category, exemplar storage provides no mechanism that would produce estimates that are more extreme than the stimulus values presented for a category. Yet perceivers in Studies 1 through 3 claimed that stimuli that were well outside the region of space that corresponded to the target category (over 4.5 standard deviations from the mean) were actually typical of the category! A parallel argument holds for the idea that the implicit system maps category responses onto subregions of perceptual space. Ashby and colleagues (Ashby et al., 1998; Ashby & Casale, 2003) argued that as a stimulus is presented and a reward follows, dopamine release increases long-term potentiation in the medium spiny cells of the caudate nucleus (Ashby & Casale, 2003). The reward occurs only

when an accurate response is made to a presented stimulus. Therefore, there should be no mapping between the category label and regions of the stimulus space where no stimuli were ever seen. Hence, assessment of the category as having attributes more extreme than those presented for the category seems at odds with this type of representation.

The present results suggest that perceivers describe stimuli as typical of the category if the stimuli fall on the correct side of the decision bound. This holds true even for stimuli that are quite unlike any of the target stimuli presented during training. The representation that seems to best fit with this result is that a decision bound is stored rather than exemplars or regions of the stimulus space. Thus, the present results support the idea that the implicit system stores a categorization rule, albeit a nonverbal one (i.e., a mathematical discriminant function). As with rule-based learning in the prefrontal cortex, the rule provides little detail regarding the specifics of the attributes of the category members.

Limitations

Studies 1 through 3 had two properties that limited their conclusiveness regarding representation. Consequently, the present data are only suggestive of a rule-based representation for the implicit neural system. First, although the category structures of Studies 1 through 3 mimic those reviewed in Ashby and Casale (2003) as requiring information integration, the fact that both dimensions were depicted as bar heights in these studies leaves open the possibility that perceivers used a verbal rule to optimize classification accuracy (e.g., "If the intelligence bar is more than 55 pixels taller than the friendliness bar, respond B"). Second, the optimal conjoint responder was able to achieve accuracies that exceeded the learning criterion.

It is important to note, however, that the main conclusions regarding between-categories contrast and within-category assimilation are not challenged by these limitations. Like the optimal classifier, the optimal verbal rule and the optimal conjoint classifier would systematically misclassify those stimuli most similar to the comparison category. Thus, biased assessments of category attributes would still be expected if perceivers used these alternative categorization strategies.

Future research should use (a) stimuli that vary on undeniably separable dimensions and (b) stimuli for which a conjoint rule cannot provide a feasible alternative explanation. Such studies could more accurately determine whether perceivers assess highly unusual stimuli as belonging to the category as long as they fit the rule defined by the optimal discriminant function. Such studies would help to clarify the representational underpinnings of categories that require information integration to optimize classification accuracy.

In addition, it could be argued that in the real world, perceivers trying to accurately classify every individual could search for additional attributes to delineate their group membership. It is true that if all of the targets are correctly classified, there would be no misclassification-induced bias. That is, if you gave an optimal classifier information that uniquely identified each individual, that person would be able to accurately classify them all. Certainly idiosyncratic, distinctive information abounds in human targets. So, for example, Seth may be a straight guy who is similar to gay men in many ways, but he has a unique mole on his left lower chin

that could serve my ability to see him as an exception within my classification scheme. However, even if such information were available, real constraints on memory and attention could limit our ability as perceivers to effectively maintain and utilize that information. It is interesting to note that studies on rule-plus-exception classification are often conducted using small numbers of stimuli and three to six discrete attribute dimensions (e.g., RULEX; Nosofsky, Palmeri, & McKinley, 1994). Whether perceivers can keep in mind many different types of exceptions to a classification scheme when the attributes are numerous and continuous and the number of stimuli is high is a question worthy of further research.

Summary

The present work makes several contributions to our understanding of category learning. We argued that stimuli that are systematically misclassified play a lesser role in assessments of group characteristics (see also Rothbart & John, 1985). Four studies altered the portion of the target category that was misclassified by the optimal classifier. As predicted, changes in the optimally misclassified stimuli led to corresponding changes in assessments of category attributes. Across studies, as the portion of the target category that was optimally misclassified increased, the between-categories contrast and within-category assimilation increased. The fact that these effects were both simultaneous and large is noteworthy because a number of researchers have failed to produce both effects simultaneously, despite Tajfel and Wilkes's (1963) original expectations that these effects would co-occur (Goldstone, 1994; Livingston et al., 1998; Tajfel & Wilkes, 1963).

It is important to note that perceivers' assessments of category attributes were quite inaccurate despite—or maybe more precisely because of—highly accurate categorization responses. Learning that optimizes classification accuracy does not necessarily provide accurate knowledge about the categories more generally. Thus, what people learn by optimizing the ability to assign items to classes may be less than optimal in guiding interactions with members of those classes.

References

- Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.
- Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jimenez (Ed.), *Attention and implicit learning. Advances in consciousness research* (Vol. 48, pp. 109–141). Amsterdam, the Netherlands: Benjamins.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 33–53.
- Bless, H., & Schwarz, N. (1998). Context effects in political judgment: Assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology*, *28*, 159–172.
- Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, *41*, 656–670.
- 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.
- Corneille, O., Klein, O., Lambert, S., & Judd, C. M. (2002). On the role of familiarity with units of measurement in categorical accentuation: Tajfel and Wilkes (1963) revisited and replicated. *Psychological Science*, *13*, 380–383.
- Eiser, J. R. (1971). Enhancement of contrast in the absolute judgment of attitude statements. *Journal of Personality and Social Psychology*, *17*, 1–10.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178–200.
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, *6*, 298–304.
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*, 608–628.
- Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27–43.
- Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, *39*, 832–845.
- 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., & 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.
- Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*, 165–188.
- Livingston, K. R., & Andrews, J. K. (2005). Evidence for an age-independent process in category learning. *Developmental Science*, *8*, 319–325.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 732–753.
- Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 288–301.
- Maddox, W. T., & Bohil, C. J. (2003). A theoretical framework for understanding the simultaneous base-rate and payoff manipulations on decision criterion learning in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 307–320.
- McGarty, C., & Penny, R. E. (1988). Categorization, accentuation and social judgement. *British Journal of Social Psychology*, *27*, 147–157.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*, 242–279.
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*, 103–122.
- Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*, 472–489.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 700–708.
- Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

- Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, *5*, 345–369.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.
- Pickett, C. L., & Brewer, M. B. (2001). Assimilation and differentiation needs as motivational determinants of perceived in-group and out-group homogeneity. *Journal of Experimental Social Psychology*, *37*, 341–348.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.
- Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues*, *41*(3), 81–104.
- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology*, *54*, 101–114.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- Tversky, A., & Kahneman, D. (1974, September). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1130.
- Whitaker, D., Bradley, A., Barrett, B. T., & McGraw, P. V. (2002). Isolation of stimulus characteristics contribution to Weber's law for position. *Vision Research*, *42*, 1137–1148.

Received May 11, 2005

Revision received November 24, 2005

Accepted December 5, 2005 ■