Social Cognition, Vol. 26, No. 1, 2008, pp. 66-101

A DECISION BOUND CATEGORIZATION APPROACH TO THE STUDY OF SUBTYPING OF ATYPICAL GROUP MEMBERS

Sarah Queller and Winter Mason
Indiana University

In subtyping, underutilization of counter–stereotypic information is argued to occur because highly atypical group members are relegated to a distinct subcategory that has minimal association with the overall group. The present work (1) avoids explicit demand to create a separate subcategory, (2) requires that perceivers overcome considerable evidence of group membership in order to subtype, (3) uses formal categorization modeling to test if a subcategory of atypical members was formed, (4) introduces a novel measure of subtyping, and (5) uses formal categorization theory to predict when subtyping should occur. The results indicate that once a subcategory of a larger group is recognized, it is included in assessments of the overall group, a finding consistent with subgrouping and not with subtyping.

Does encountering information that challenges our stereotypes cause us to unlearn or revise our stereotypes? The intuitive answer to this question is "yes." Yet, research has shown that our stereotypes are often resistant to change. In explaining these findings, researchers have proposed a variety of mechanisms that allow stereotypes to persist despite exposure to counter–stereotypic group members. For example, stereotype–inconsistent group members can be discounted as anomalous, stereotype–disconfirming evidence can be ignored or reinterpreted to fit with the stereotype, or stereotype–disconfirming group members can be subtyped (Darley & Gross, 1983; Hamilton & Rose, 1980; Hewstone, Hopkins, & Routh, 1992; Hewstone, Macrae, Griffith, & Milne, 1994; Johnston & Hewstone, 1992; Razran, 1950; Rothbart & John, 1985; Sagar & Schofield, 1980; Taylor, 1981; Weber & Crocker, 1983; Wilder, 1984).

Sarah Queller and Winter Mason, Department of Psychology, Indiana University.

This research was supported in part by a grant from the National Science Foundation and was supported jointly by Indiana University and University of California, Santa Barbara.

We would like to thank Greg Ashby, Diane Mackie, Dave Hamilton, Stan Klein, Terry Schell, and Jim Sherman for their contributions to this project.

Correspondence concerning this article should be addressed to Sarah Queller, Department of Psychology, Indiana University, Bloomington, IN 47405. E–mail: queller@indiana.edu.

Subtyping is the focus of the present article. When highly atypical group members who differ from stereotypic expectations in similar ways are encountered, the atypical group members can be subtyped (Brewer, Dull, & Lai, 1981; Taylor, 1981) or relegated to a separate subcategory that is mentally isolated from the overall social group (Johnston & Hewstone, 1992; Queller, Mackie, & Stroessner, 1996; Weber & Crocker, 1983; see Richards & Hewstone, 2001 for a review). In this manner, the stereotype of the overall social group can be maintained, despite knowledge that a number of members violate stereotypic expectations. Thus, for example, black businessmen may not affect stereotypes of blacks in general (Devine & Baker, 1991).

Note that subtyping is different from simple discounting of atypical group members precisely because it posits the formation of a separate subcategory of atypical group members. When group members are simply discounted, they are not treated as a separate subcategory. Rather, they remain part of the overall category but the more atypical they are, the less they contribute to group assessment relative to typical members (Anderson, 1980; Posner, 1974, as cited in Rothbart and John, 1985). Decreased impact of the atypical members comes about because they are not retrieved as good instances of the category. For example, ostriches have a limited impact on the assessment of the category birds because they are not readily retrieved when thinking of birds.

In contrast to discounting, Maurer, Park, & Rothbart (1995) describe subtyping as forming a subcategory of highly atypical group members that does not remain under the umbrella of the stereotyped group. Effectively, the highly atypical group members are recategorized and no longer retain an association with the stereotyped group. This loss of association with the stereotyped group is responsible for the lack of impact that subtyped group members have on assessments of the overall group. In contrast, Maurer and her colleagues describe a different scenario when the atypical group members are only moderately atypical. In this case, they claim a subcategory of atypical group members is formed that does remain under the umbrella of the larger stereotyped group. They call this subgrouping as opposed to subtyping. With subgrouping, the retained association between the subgroup and the larger stereotyped group leads to stereotype change in the form of increased perceptions of group variability.[1]

This paper further investigates the phenomenon of subtyping. First, we review the current state of evidence supporting subtyping as a mechanism of stereotype maintenance and we raise questions about this evidence. In the service of providing a novel approach to studying the categorization aspects of subtyping, we next review the Decision Bound Model of categorization (Ashby & Gott, 1988) and describe relevant data collection and analysis techniques. Finally, we report two studies investigating subtyping from the perspective of the Decision Bound Theory of categorization. The results of these studies have important implications for understanding the process of subtyping.

---

1. It is important to make the distinction between the definitions of subtyping and subgrouping described above because some research has referred to both of these phenomena as subtyping. Some of the earliest research, however (e.g., Brewer et al., 1981; Talyor, 1981) acknowledges that highly atypical subtypes might be minimally or not at all associated with the overall group. Effectively, subtyping now refers to subgroups that have lost their association with the larger category.

EVIDENCE THAT SUBTYPING ALLOWS STEREOTYPE MAINTENANCE

Subtyping that allows stereotype maintenance has been repeatedly described in the context of the "concentrated/dispersed" paradigm. In this paradigm, participants read statements describing members of a stereotyped group. The descriptions contain stereotype–consistent, stereotype–inconsistent, and stereotype–irrelevant statements. Each group member is described by several statements. In the concentrated condition, all of the stereotype–inconsistent statements are attributed to a few highly atypical group members. In the dispersed condition, the same stereotype–inconsistent statements are attributed to a larger number of atypical group members who are, consequently, only moderately atypical. The main dependent measures are stereotypic and counter–stereotypic trait ratings. A number of researchers using this paradigm have found that the group is rated less stereotypically when the stereotype discrepant information is attributed to a number of moderately atypical group members than when it is attributed to only a few, highly atypical group members (for a review, see Richards & Hewstone, 2001; for exceptions, see Gurwitz & Dodge, 1977; Hewstone, Johnston, & Aird, 1992).

Of course, although trait ratings may well indicate that the stereotype has been maintained, these ratings alone provide no evidence for the categorization aspect of subtyping. To address this issue, researchers have explored a number of creative measures in search of evidence that shows the highly atypical group members were categorized separately from the rest of the stereotyped group. The most common technique in these studies is to have participants sort the descriptions of the individual group members so that members similar to one another remain in the same pile and members in each pile are somehow different from those in other piles.

Either counting the number of piles (Weber & Crocker, 1983) or performing hierarchical cluster analyses on the sorted piles (Johnston & Hewstone, 1992) lends support to the idea that the highly atypical group members (concentrated condition) form a coherent mental cluster that is separated from the larger group. In contrast, the moderately atypical group members (dispersed condition) do not form a coherent cluster. The acknowledged problem with this measure (Hewstone et al., 1994; Johnston & Hewstone, 1992; Queller et al., 1996) is that it presents an experimental demand to create distinct subcategories within the overall group. Just because people can mentally segregate the highly atypical group members when asked to perform a sort task does not mean that they spontaneously do so when no prompt to sort is present. The sort task's legitimacy as a measure of subtyping is further weakened by the findings of Hewstone, Johnston, & Aird (1992). For a homogeneous group, they found a reversal of the typical finding, with more stereotype change in the concentrated condition than in the dispersed condition. This was despite the fact that the sorts indicated that clustering of the highly atypical group members in the concentrated condition was high, as if the highly atypical group members had been subtyped. Given these results and the experimental demand associated with the sort task, it seems that the sort task presents only weak evidence for the recategorization aspect of subtyping.

In another attempt to provide support for the recategorization aspect of subtyping, Johnston and Hewstone (1992) collected reading time data and showed that participants spent more time reading about the moderately atypical group members in the dispersed condition than about other group members (i.e., highly atypical members in the concentrated condition or typical members in either condition). They suggested that participants who encountered moderately atypical group members engaged in inconsistency resolution but those who encountered highly atypical group members did not (Hastie & Kumar, 1979). Although this finding may suggest that participants dismissed the highly atypical group members as not belonging to the overall group and therefore not requiring inconsistency resolution, it does not indicate whether the mechanism of dismissal was related to discounting (i.e., an atypical member simply does not cue the category very well because they have fewer category–typical features) or was related to subtyping. Thus, the reading time measure does not directly address the recategorization aspect of subtyping.

A final measure used as evidence of subtyping is clustered recall (Hewstone et al., 1994; Queller et al., 1996). Research in the cognitive literature suggests that if participants study a list consisting of items belonging to different categories, subsequent recall of the items on the list is clustered in terms of category membership (Roenker, Thompson, & Brown, 1971). Extending this idea to the realm of subtyping, if the highly atypical group members are thought about as a separate category, behaviors performed by highly atypical group members should tend to be recalled in sequence and behaviors performed by typical group members should tend to be recalled in sequence. In contrast, behaviors performed by moderately atypical group members should be recalled in an order intermixed with the behaviors performed by typical group members. Results support these claims, showing more clustering of recall for behaviors performed by highly atypical group members in the concentrated condition than for those performed by mildly atypical group members in the dispersed condition. These results suggest that, even in the absence of the demands of a sort task, participants mentally segregate the highly atypical group members into a separate category.

Queller and colleagues (1996), however, raised an alternate interpretation of these findings. Because the highly atypical group members in the concentrated condition are described by only stereotype–inconsistent behaviors, a high degree of clustering in this condition may simply be a result of clustering by behavior type at recall. That is, the categories responsible for producing clustered recall may be the categories of stereotype–consistent *behaviors* versus stereotype–inconsistent *behaviors* and not the categories of typical versus atypical *group members*. Hewstone and colleagues (1994) pointed out that recalling behaviors performed by different *individuals* in a clustered manner (person–based clustering) would also lead to elevated clustering in terms of typical and atypical group members in the concentrated condition compared to the dispersed condition.

Using clustered recall as a measure of subtyping in the concentrated/dispersed paradigm thus requires that at least two alternate interpretations be eliminated before clustering in terms of typical and atypical group members can be shown. Unfortunately, elimination of those hypotheses relies on failure to reject the null hypotheses that no behavior–based or individual–based clustering occurred. Failure to reject these null hypotheses requires more strict power analyses than have been done to date. In addition, these competing hypotheses can only be tested in the dispersed condition, where behavior type and individual are not confounded with group member typicality. Given these constraints, clustered recall within the context of a concentrated/dispersed paradigm may not be a straightforward indicator of categorization.

In summary, several attempts have been made to show that the recategorization of highly atypical group members into a subtype can lead to stereotype mainte-

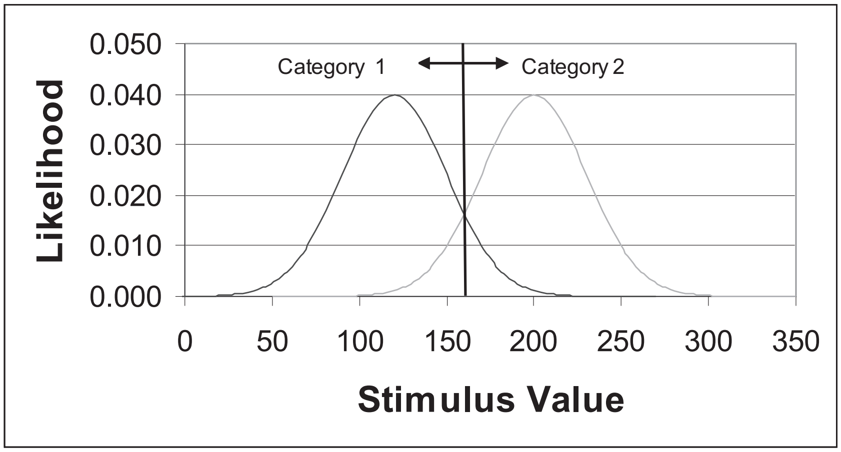

FIGURE 1. Optimal decision bound between two categories defined in one dimension. The bound is actually a single point at x = 200. It is depicted here as a vertical line simply to facilitate viewing.

nance. Although the trait ratings suggest that highly atypical group members may be ignored when assessing the overall group, the data are much weaker in providing support for the contention that the formation of a separate mental category of atypical individuals is responsible for this effect.

THE DECISION BOUND THEORY OF CATEGORIZATION

It appears that novel techniques may be required to provide additional evidence of whether or not highly atypical group members are relegated to a separate category. The present research relies on a popular model of cognitive categorization, Ashby's Decision Bound Theory of Categorization (Ashby, 1992a; Ashby & Gott, 1988; Ashby & Maddox, 1992) to suggest such novel techniques.

Decision Bound Theory is based on the idea that categorization can be described by the determination of the boundaries that best separate one category from another on the basis of relative likelihood. For two categories of equal size, an observer assigns a new individual to the category that is more likely, given that individual's characteristics. For example, I might class a person who is 5′2″ tall as a woman because the likelihood of a woman being is 5′2″ is higher than the likelihood of a man being 5′2." Observers guess at category membership when a new target's characteristics are equally likely to have come from either category. Thus, the optimal decision bound between two categories of equal size is defined by the set of points where the probability density functions (pdf's) of the two categories intersect (see Appendix A for details.) For a one–dimensional space, the decision bound can be a single point along the dimension (see Figure 1).

When category sizes are not equal, the decision rule adjusts for the relative frequencies of the categories, enlarging the region corresponding to the more frequent category. For example, suppose Category 2 members are more frequent in occurrence than are Category 1 members in Figure 1. In this case, the optimal bound that best distinguishes Category 1 members from Category 2 members shifts toward the Category 1 mean, a point where the Category 1 pdf is lower than the Category 2 pdf (see Appendix A).

It is important to realize that the optimal decision bound does not necessarily result in 100% categorization accuracy. An observer could conceivably learn to distinguish items in different categories with 100% accuracy if he or she accurately perceived and represented each stimulus in the perceptual space and if he or she accurately recalled the group membership associated with each stimulus. This type of accurate learning is possible even when categories are highly overlapping if perceptual noise is low, feedback is consistent and accurate, the number of stimuli is small, and no stimuli with identical characteristics have different category memberships. However, when these conditions are not met, an alternative strategy is to optimize the number of correct categorizations by assessing relative likelihood and responding with the category that is most likely given the characteristics of the stimulus. Although this strategy does not produce perfectly accurate category responses, it produces the decision bound that is optimal in the sense that it maximizes the number of correct responses.

An optimizing strategy such as that proposed by Decision Bound Theory is highly useful when judging the category membership of novel stimuli not encountered during training. When encountering a novel stimulus, the observer must decide category membership without the benefit of prior feedback regarding that particular stimulus. In this situation, an optimizing strategy such as relying on relative likelihood from past experience will result in maximum accuracy.

In a paradigm often used to study categorization performance (Ashby & Gott, 1988; Ashby & Maddox, 1992; McKinley & Nosofsky, 1995), observers encounter novel stimuli, one at a time, and decide to which of two categories each stimulus belongs. After each categorization attempt, observers get feedback telling them whether their categorization was correct or incorrect. Through numerous trials with feedback, observers learn about the categories.

In the context of this paradigm, participants' responses can be analyzed to determine the likelihood of a model given the data. For a given model type, the parameters of that model can be varied until the best fit between the model and the data is achieved. The comparative fit of two different nested model types can be compared to see which better accounts for a respondent's categorization responses by using a Likelihood Ratio Test (Ashby, 1992b, see Appendix A).

In summary, Decision Bound Theory views categories as regions in the stimulus space that are separated by decision bounds. Comparing fits of competing decision bound models can pinpoint the decision bound that best describes an individual's categorization responses. Such analyses of model fit have shown that, with feedback during category learning, observers approach optimality in their categorizations.

USING DECISION BOUND THEORY TO STUDY SUBTYPING

The cognitive categorization paradigm used in the present studies (and in the cognitive experiments described above) differs from paradigms traditionally used in social psychology in several ways. The foremost difference is that participants learned about artificial categories. This has advantages and weaknesses. By using hundreds of targets, the experimenter can control the properties of the targets and the distributional properties of the groups. Tightly controlled stimuli allow for precise modeling of the processes underlying categorization and subtyping. Unfortunately, this also requires the use of unrealistic stimuli, so that people are establishing newly learned stereotypes about simplistic targets using accurate, objective information that may not generalize to the complex process of forming stereotypes and subtypes about existing groups. The present paradigm does provide an advantage in that it assesses whether atypicality on a single dimension is sufficient to cause people to create a separate mental subcategory that is excluded from assessments of the overall group. Although Hewstone and colleagues suggest that typicality is the key variable in subtyping (Johnston & Hewstone, 1992; Richards & Hewstone, 2001), this claim has not been strictly tested using minimal conditions to isolate that single variable.

A second difference is that participants' data are compared to what would be expected given a hypothesized Decision Bound Model. Thus, none of the studies reported here pit one experimental condition against another. Rather, they pit data against a number of proposed models with a resultant determination of which model best describes participants' performance.

Probably the largest departure from what is typically done in social psychological experiments is the use of single subject analyses. That is, the hypothesized models are fit for each subject separately. In a sense, this is like treating each subject as a separate experiment, and multiple subjects as replications of the experiment. Of course, individual differences are built into the "replications." This type of analysis is appropriate when you need to know the strategy that a particular individual followed in order to interpret their subsequent results. In addition, averaging across observers might obscure the use of different strategies by different individuals (Estes, 1956). Finally, the increased symmetry of averaged data (e.g., Central Limit Theorem) gives unfair advantage to models that assume symmetry in the data (Ashby, Maddox, & Lee, 1994). Consequently, model comparisons are often done at the individual level.

In a final departure from traditional social psychology paradigms, the number of subjects run in the types of categorization experiments described above as well as in the present studies is typically quite small (4–8), and the analysis of each subjects' many data points is intensive.

Aside from these differences, the application of Decision Bound Theory to the study of subtyping requires more than direct use of the classic two–category classification task with corrective feedback. The present studies involve a series of tasks that allow us to first establish whether participants recognize a subcategory of highly atypical Group A members—in the absence of explicit demands to do so—and then provide a novel test of whether that atypical subcategory is subsequently ignored when making judgments about the overall group.

STUDY 1

The goal of Study 1 is similar to that of other studies that investigate subtyping as a means of stereotype maintenance. That is, an attempt is made to provide evidence that perceivers recategorize the atypical group members such that these members do not affect subsequent judgments about the overall group. In accord with this goal, this study (1) encourages participants to recognize an atypical subtype in the absence of any experimental demand to separate the group into subcategories; (2) reflects the idea that any recategorization of atypical group members involves overriding information suggesting the atypical members, in fact, belong to the group; (3) provides a novel means of evaluating whether the atypical group members are perceived as a subcategory that is distinct from the subcategory of typical group members; and (4) introduces a categorization measure (along with traditional measures of central tendency and dispersion) to assess the mental inclusion or exclusion of the atypical subcategory in assessments of the overall group.

Study 1 proceeded as follows: Participants first learned to distinguish both typical and atypical Group A members from uniformly distributed individuals who were not members of Group A. The values of the stimuli presented are depicted in Figure 2 (top panel). During this A/Not A learning phase, a participant viewed the scholastic ability score of a target person and attempted to decide whether that person was a member of Group A or not. After each attempt, the computer provided corrective feedback. This continued until the participant had categorized each of the stimuli one time.

The Group A members were bimodally distributed with a major mode corresponding to the average typical group member and a minor mode corresponding to the average atypical group member. Thus, the atypical group members were atypical on the basis of (1) much lower frequency compared to the typical group members, and (2) values of scholastic ability that were far removed from those of the typical group members. The individuals who were not members of Group A were uniformly distributed throughout the stimulus space. That is, the nonmembers exhibited a broad variety of scholastic abilities with no single range of values being particularly diagnostic without referring to the abilities of the Group A members. Overlaying these member and nonmember distributions reveals that nonmembers are more likely given very low values of scholastic ability. Group A members (the typical ones) are more likely than nonmembers given moderately low scholastic ability. Nonmembers are again more likely than Group A members for intermediate scholastic ability. Group A members are more likely than nonmembers for moderately high scholastic ability. Finally, for very high values of scholastic ability, nonmembers are more likely than Group A members. Thus, on the basis of relative likelihood, participants were encouraged to see the atypical group members as a distinct subcategory of Group A.

To see whether participants viewed the atypical group members as a distinct subcategory or, alternately, simply discounted them as not belonging to Group A, two competing models were fit to each participant's data: First, a NotA | A | NotA model with two decision bounds and one continuous Group A category and, second, a NotA | A | NotA | A | NotA model with four decision bounds and two Group A subcategories. (The " | " indicates the position of a decision bound.) A better fit for the two bound model with the higher bound excluding the atypical Group A members would provide evidence that the atypical group members were discounted and seen as not belonging to Group A but would provide no evidence that the atypical group members were seen as a coherent category. A better fit for the four bound model would indicate that the atypical group members were seen as a subcategory that was distinct from the subcategory of typical group members.

Showing that two distinct subcategories of Group A members were formed is not sufficient to establish subtyping. To establish subtyping, the atypical group members must not only form a separate subcategory, they must also have minimal impact on assessments of the overall group. For example, aggressive business women, who are surely women after all, can be relegated to a subcategory that is separate from other women. As a consequence, the aggressive business women will have little impact on assessments of what women, in general, are like. Decision Bound The-

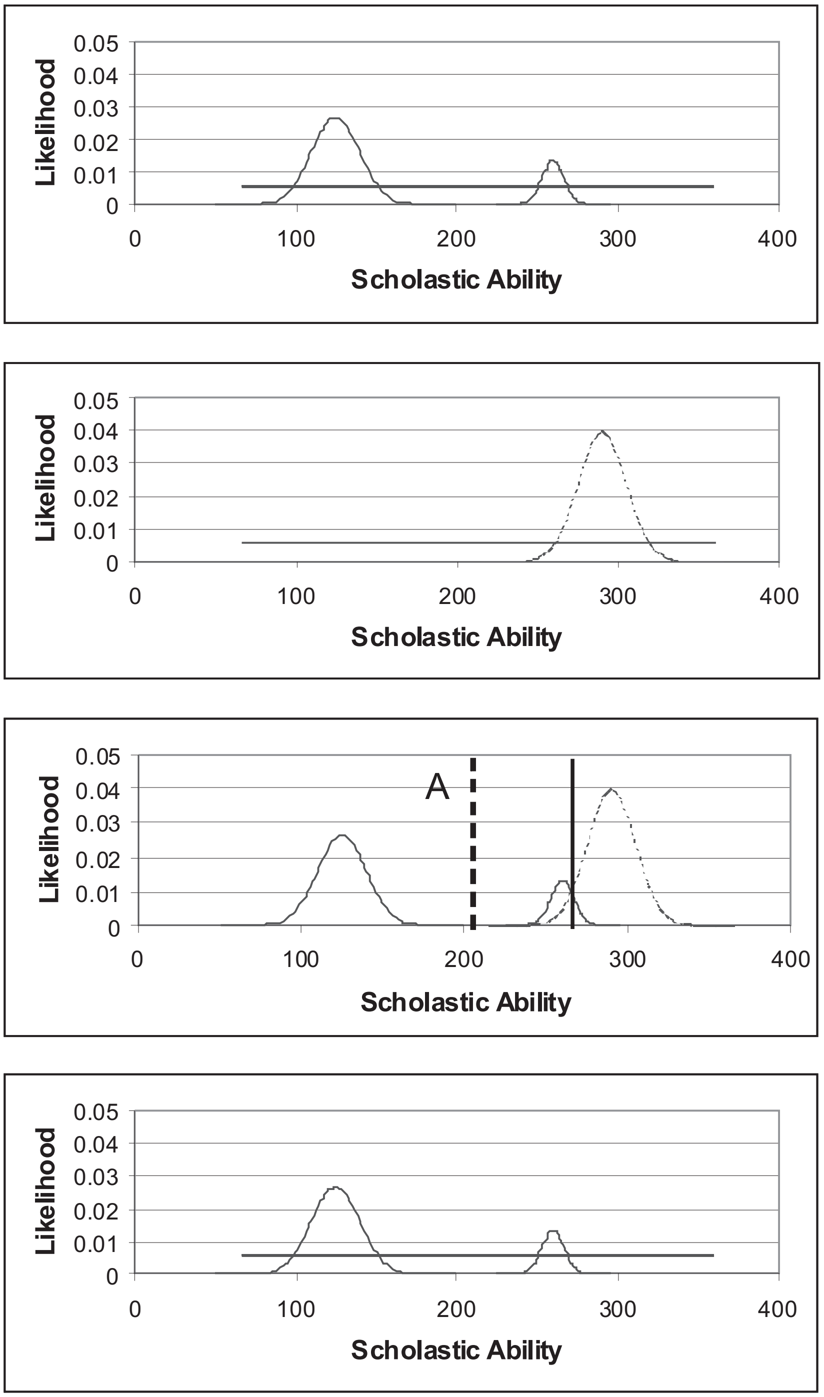

FIGURE 2. Probability density functions for Study 1. Solid function = Group A pdf; Dotted function = Group B pdf; Dashed function = Not A pdf or Not B pdf. The solid vertical line in the third panel indicates the optimal bound. The dotted vertical line indicates the subtyping bound.

ory suggests a novel measure of whether or not the atypical group members are included in assessments of the overall group. Specifically, would those participants who created a distinct subcategory for the atypical Group A members later exclude those members in the context of a Group A versus Group B categorization task? If so, this would provide evidence of subtyping leading to stereotype maintenance.

In order to perform in a Group A versus Group B categorization task, participants first had to learn about Group B members. They did this in the B/Not B learning phase. The B/Not B learning phase was similar to the A/Not A learning phase except that participants learned about Group B members versus individuals who were not members of Group B (see Figure 2, second panel). Feedback was provided after each categorization attempt. Note that the Group B members' scholastic abilities were similar to those of the atypical Group A members.

To see if the atypical Group A members were included in assessments of Group A in the context of a Group A versus Group B categorization task, participants completed an A/B test phase. In the A/B test phase, participants were presented with novel stimuli and decided whether each stimulus belonged to Group A or to Group B. Participants did not receive feedback during the A/B test phase. Figure 2, third panel, depicts the distributions of Group A versus Group B members presented during the A/B test phase. Because equal numbers of Group A and Group B members were presented during learning, the optimal decision bound is at the point of intersection between the Group A pdf and the Group B pdf. This optimal bound assumes that the atypical Group A members are included in assessments of Group A. If the atypical Group A members are *not* included in assessments of Group A, the "subtyping bound" would be placed in the space between the typical Group A members and the atypical Group A members. Note that the optimal bound is higher than the mean of the atypical Group A members whereas the subtyping bound is lower than the mean of the atypical Group A members (see Figure 2, third panel). Thus, the positioning of the decision bound in the A/B categorization task should indicate whether the atypical Group A members were included in, or excluded from, assessments of Group A.

Suppose the atypical Group A members were excluded when making Group A versus Group B categorization judgments in the A/B test phase. This would suggest that subtyping had occurred. However, an alternate possibility is that the atypical Group A members were reinterpreted as belonging to Group B after the B/Not B learning phase. After all, the Group B members were learned about more recently and were more numerous than the atypical Group A members. This may have led participants to question their original classification of the atypical Group A members as belonging to Group A. To eliminate the hypothesis that the atypical Group A members were reinterpreted as belonging to Group B, a second test phase was included where participants made Group A versus Not Group A judgments in the absence of feedback. In this A/Not A test phase, participants were presented with novel stimuli with the same distributional properties used in the A/Not A learning phase. Participants classified each stimulus as belonging to Group A or as not belonging to Group A in the absence of feedback. If the atypical Group A subcategory was accounted for in the A/Not A test phase, this would argue against the idea that the atypical Group A members had been reinterpreted as belonging to Group B.

The order of the A/B test phase and the A/Not A test phase was counterbalanced. This was done because of a concern that previous performance of the A/Not A test phase might affect performance in the A/B test phase and vice versa.

TABLE 1. Proposed results that would support subtyping and the exclusion of atypical group members in assessing the overall group

Finding	Implication
A/Not A Learning Phase	
The four bound NotA|A|NotA|A|NotA model fits better than the two bound NotA|A|NotA model	Group A was accurately learned and a sub–category of atypical group members was recognized
B/Not B Learning Phase	
A two bound NotB|B|NotB model fits reasonably well	Group B was accurately learned
A/B Test Phase	
Subtyping bound fits better than optimal bound	The atypical Group A members were not included as belonging to Group A
A/Not A Test Phase	
The four bound NotA|A|NotA|A|NotA model fits better than the two bound NotA|A|NotA model	The atypical Group A members were not simply re–interpreted as belonging to Group B
Central Tendency	
Estimate is near the mean of the typical Group A members	atypicals were not included as belonging to Group A OR participant reported an accurate modal value
High	
High value does not include values representative of atypical Group A members	The atypical Group A members were not included as belonging to Group A
Low	
No critical predictions relevant to subtyping	

Following both the A/B and the A/Not A test phases, participants also performed the more traditional estimates of central tendency. A measure of dispersion was also included. Note, however, that with stimuli that include highly atypical group members, a measure of central tendency can be difficult to interpret. Suppose you were asked to estimate the average scholastic ability for members of Group A. If you computed the mathematical average of all Group A members, the value would be 170. However, suppose you thought “average” meant “typical.” In this case you might instead report the value of the major mode of the Group A scholastic abilities, or 125. Either of these responses would be accurate and simply reflect different interpretations of what the experimenter requested. The latter response, however, would not change depending on whether the atypical group members were included or excluded in the assessment of central tendency. Thus, if a participant reports a central tendency that is coincident with the mean of the typical Group A members, this does not necessarily suggest that the atypical Group A members have been ignored. Despite these problems of interpretation, a measure of central tendency was included in the present study for comparison to previous work.

Given the previous discussion of the effects of nonsymmetric distributions on estimates of central tendency, it seemed important to choose a measure of dispersion that would encourage reporting of the most extreme value of scholastic ability that was considered as belonging to Group A. Thus, participants were asked to report the highest and lowest values of scholastic ability that described Group A members (Park & Judd, 1990). Using the combination of techniques and measures described here, it should be possible to determine whether participants created a separate category for the atypical Group A members that did not affect judgments about the overall group. Table 1 shows the pattern of results that would support subtyping of the atypical Group A members

METHOD

Participants

Eight participants (4 female and 4 male) were recruited from the University of California, Santa Barbara (UCSB) student population and were paid $18 for their participation in the experiment.

Stimuli

Scholastic ability scores for each individual were displayed in bar form by varying the height of a bar on a computer screen. Stimulus values are described in computer screen pixel units. All stimulus values were rounded to the nearest whole number.

In the A/Not A learning phase, the scholastic ability scores for 60 typical Group A members were drawn from a normal distribution with mean $\mu = 125$ and variance $\sigma^2 = 225$ (see Ashby, 1992b, for the random sampling technique used). The scores for 30 atypical Group A members were drawn from a normal distribution with mean $\mu = 260$ and variance $\sigma^2 = 49$. Thus, twice as many typical Group A members were presented as atypical Group A members. The distribution of the atypical Group A members was well removed from the distribution of the typical Group A members such that there was essentially 0% overlap between these two distributions. The scores for the 148 individuals who were not members of Group A were uniformly distributed across the stimulus space and thus provided "background" noise from which the Group A members had to be distinguished. Note that the scores of the individuals who were not members of Group A were evenly spaced throughout the stimulus region and ranged from 66 to 360 in increments of 2. This procedure for generating stimuli was repeated five times to create five blocks of 238 trials per block (60 typical and 30 atypical Group A members plus 148 nonmembers).

In the B/Not B learning phase, scores for 90 Group B members were drawn from a normal distribution with mean $\mu = 290$ and variance equal to that of the typical Group A members ($\sigma^2 = 225$). The number of Group B members thus equaled the total number of Group A members presented in the A/Not A learning phase. There was no overlap between the distribution of the typical Group A members and the Group B members. The overlap between the (atypical) Group A members and the Group B members was 5.8%. The scores for the 148 individuals who were not members of Group B were the same uniformly distributed scores used to describe the individuals who were not members of Group A in the A/Not A learning phase. Again, the procedure was repeated five times to create five blocks of 238 trials per block.

In the A/B test phase, 60 stimuli were drawn from the typical Group A distribution ($\mu = 125$, $\sigma^2 = 225$), 30 were drawn from the atypical Group A distribution ($\mu = 260$, $\sigma^2 = 49$), and 90 were drawn from the Group B distribution ($\mu = 290$, $\sigma^2 = 225$). This procedure was repeated three times to generate three blocks, each containing 180 stimuli.

In the A/Not A test phase, 60 stimuli were drawn from the typical Group A distribution ($\mu = 125$, $\sigma^2 = 225$), 30 were drawn from the atypical Group A distribution ($\mu =$

260, $\sigma^2 = 49$), and 148 stimuli were the same uniformly distributed stimuli used in the learning phases. This procedure was repeated three times to generate three blocks, each containing 238 stimuli.

Throughout Study 1, each participant saw the exact same stimuli. However, the order of presentation within each block was randomized, so that each participant viewed the stimuli in a different order.

Procedure

Participants were introduced to the study as follows: (During the past two years, the University of Michigan's Institute for Survey Research has been conducting a survey of fifteen different communities. In this study, a variety of psychological measures were collected from members of different social groups. These groups consisted of groups defined by social roles, such as doctor or engineer, as well as gender groups, ethnic groups, and socioeconomic status groups. This survey is essentially like a census database that can be broken down in a variety of ways. In addition to collecting data from the group members themselves, we are also studying how these groups are perceived by others like yourself. In this study, you will be presented with information about the scholastic abilities of some of the groups that were surveyed. Scholastic ability was measured using a combination of performance measures and behavioral measures. We will not tell you which real life group you are learning about for two reasons. First, we want to maintain confidentiality regarding the survey respondents. Second, and more importantly for this study, we want your perceptions of each group under conditions where you do not have any knowledge of exactly which group you are learning about. Therefore, as you learn about the different groups, the groups will be referred to with a letter designation. For example, you will first be learning about a group we will refer to as Group A.)

Participants then read instructions telling them they would be asked to view a succession of scholastic ability scores (represented as bar heights) and would make decisions about whether each score was that of a Group A member or that of an individual who was not a member of Group A. The experimenter then informed them (1) that the individuals who were not members of Group A consisted of individuals from a mixture of other groups; (2) that because the scholastic ability scores were from real groups, there was some overlap between groups and they would not be able to achieve 100% accuracy; (3) that presentation of group members' scores would be divided into five blocks and that they would get to take breaks between blocks; (4) that the breaks were simply inserted for their own comfort, and that they would be learning about the same group(s) throughout all five blocks; and (5) that the scores of the surveyed individuals would appear in random order and that looking for order effects (e.g., every third score was that of a Group A member) would only impair their accuracy.

On each trial, the scholastic ability of a typical Group A member, an atypical Group A member, or an individual who was not a member of Group A was presented. The participant's task was to press the "z" key on the computer keyboard if they thought the individual was a member of Group A, and to press the "/" key if the individual was not a member of Group A. The participant's category response on each trial was recorded. Participants were encouraged to respond as quickly as possible while maintaining accuracy. Feedback was provided on every trial. When participants made a correct response, a pleasant, high–pitched tone sounded and green text appeared. The text told them that their category choice was correct and displayed the group membership of the previously viewed individual. When they made an incorrect response, an unpleasant, low–pitched tone sounded and yellow text appeared. The text informed them their response was incorrect and displayed the group membership of the previously viewed individual. Each trial was response terminated, and the feedback remained on the screen for 1 second before continuing on to the next trial. The delay between the removal of the feedback text and the presentation of the next stimulus was 500 ms. A participant–terminated break occurred between blocks (238 stimuli/block, 5 blocks). During the break, the computer displayed the percent correct attained by the participant in the previous block. The first learning phase took approximately one hour to complete.

The B/Not B learning phase followed after a short break. The instructions for the second phase were similar to those for the first phase, except participants were told they would be making judgments about a new group that would be referred to as Group B versus individuals who were not members of Group B. Participants were also told that the Group A members were removed from the data that would be presented in the B/Not B phase, so that they should not worry about what they learned in the first phase of the experiment while making judgments about Group B. The procedure was identical to that for the A/Not A learning phase except the Group B and not Group B stimuli were used with the "z" key assigned to designate Group B members and the "/" key assigned to designate individuals who were not members of Group B. The second learning phase also took about one hour to complete.

Each participant returned the following day to complete the two test phases of the experiment. The order of the A/B test phase and the A/Not A test phase was counterbalanced. Instructions for the test phases alerted participants that they would receive no feedback during the task and would have to rely on what they had learned previously to complete the task. In each of the test phases, three blocks were presented with participant–terminated breaks between blocks. During the test phase breaks, participants did not receive any information regarding their performance.

In the critical A/B test phase, participants viewed scholastic ability scores of both Group A and Group B members and categorized each stimulus as belonging to a Group A member ("z" key) or a Group B member ("/" key). No feedback was provided and category responses were recorded.

At the end of the third block of the A/B test phase, participants adjusted a minimum height bar using the arrow keys until the height of the bar "represent[ed] the score of the average member of Group A." When this measure of the perceived central tendency was completed, participants completed a measure of dispersion: They first adjusted a minimum height bar until it represented the lowest Group A score. When that was done, they adjusted a final minimum height bar until it represented the highest Group A score.

The A/Not A test phase was similar to the A/B test but participants once again made assessments about whether the score was that of a Group A member ("z" key) or that of an individual who was not a member of Group A ("/" key). The stimuli presented were similar to those presented in the A/Not A learning phase. Participants were explicitly told to focus only on whether the score reflected a Group A member or not and to ignore what they knew about Group B for this task. Thus, if participants had reinterpreted atypical Group A members as belonging to Group B during the B/NotB learning phase, they should respond "Not A" to the atypical Group A members in the A/NotA test phase. In contrast, if participants responded

"A" to the atypical Group A members in this test phase, this would indicate that they had not reinterpreted the atypical Group A members as belonging to Group B. No feedback was provided and category responses were recorded. After completing the third block of the A/Not A test phase, participants engaged in the same average, low, and high assessments regarding Group A that were described for the A/B test phase.

Participants were then debriefed, thanked for their participation, and dismissed.

RESULTS

Category Responses in the A/Not A Learning Phase. To determine whether two distinct subcategories of Group A members were recognized in the A/Not A learning phase, the fit of a four bound model (NotA | A | NotA | A | NotA) was compared to the fit of a two bound model (NotA | A | NotA).

These two models were compared for each participant's data in the third through fifth blocks of the A/Not A learning phase. The model fits and bound positions are displayed in Table 2. Note that this table can be easily interpreted by inserting the bound values (B1, B2, B3, B4), in order, at the positions of the ' | ' marks in the Model column. Thus the first row would be read as NotA–106–A–147–NotA–231–A 275–NotA, meaning the typical Group A members were clustered between the values of 106 and 147 whereas atypical Group A members were clustered between the values of 231 and 275.

For five of the participants, the four bound model fit significantly better than the two bound model (Participants 1, 3, 4, 5, 8). In contrast, a two bound model fit just as well as a four bound model for Participant 6 and for Participant 7's third and fourth blocks. Notice that for Participant 7's fifth block, the best fit model was a four bound model. However, a closer look at the values of the bounds indicates that the third and fourth bounds are equal (x = 265). This finding combined with a glance at the raw data show that the four bound model fits better than the two bound model because of a single point that was designated as belonging to Group A at x = 265.[2]

If you drop this one point, Participant 7's responses indicate a single cluster with an upper bound of 143. Combined with the fact that Participant 7's third and fourth block responses were consistent with a two bound model, this participant clearly ignored the atypical Group A members during the A/Not A learning phase. In addition, the values of the bounds for Participants 6 and 7 indicate that the atypical Group A members were not included as belonging to Group A. Finally, Participant 2 responded in a manner that indicated the model that best fit the data was a model with only "Not A" responses. This participant clearly did not learn the category structure and will not be discussed further.

The optimal bounds given the stimuli that were presented are also shown in Table 2 for comparison. Note that, although five participants' data were best fit with the optimal model type (i.e., a four bound model), the positioning of their bounds only approached optimality.

In summary, the model fits indicate that five of the participants recognized two

2. It should be noted that model comparisons based on maximum likelihood are sensitive to outliers. However, maximum likelihood model fits are used here because they allow a statistical comparison between the fits of different models. The ability to do such comparative tests is not available with other types of model fits (e.g., least squares).

TABLE 2. Optimal bounds for presented stimuli (PS) and best fitting models and bound positions for participants' responses in the last three blocks of the A/Not A learning phase of Study 1

Part #	Block	Best Fit Model	B1	B2	B3	B4	Error	–lnL	χ^2
PS		optimal bounds	106	144	248	272	13	94	
1	3	NotA|A|NotA|A|NotA	106	147	231	275	21	122	62**
1	4	NotA|A|NotA|A|NotA	105	150	235	266	17	109	72**
1	5	NotA|A|NotA|A|NotA	91	145	240	278	17	102	100**
3	3	NotA|A|NotA|A|NotA	73	131	252	281	29	126	32**
3	4	NotA|A|NotA|A|NotA	40	125	269	295	38	124	25**
3	5	NotA|A|NotA|A|NotA	80	132	256	297	27	132	42**
4	3	NotA|A|NotA|A|NotA	99	137	263	299	9	99	143**
4	4	NotA|A|NotA|A|NotA	94	137	250	288	11	94	148**
4	5	NotA|A|NotA|A|NotA	96	141	248	286	7	96	207**
5	3	NotA|A|NotA|A|NotA	94	151	238	288	16	95	117**
5	4	NotA|A|NotA|A|NotA	97	148	240	292	15	90	132**
5	5	NotA|A|NotA|A|NotA	82	143	241	282	15	89	123**
6	3	NotA|A|NotA	63	185	—	—	57	110	5.8, ns
6	4	NotA|A|NotA	77	159	—	—	41	89	0, ns
6	5	NotA|A|NotA	88	150	—	—	33	84	0, ns
7	3	NotA|A|NotA	88	143	—	—	11	43	0, ns
7	4	NotA|A|NotA	85	147	—	—	12	39	0, ns
7	5	NotA|A|NotA|A|NotA	88	143	265	265	6	26	61**
8	3	NotA|A|NotA|A|NotA	97	145	240	259	22	114	34**
8	4	NotA|A|NotA|A|NotA	95	139	240	274	20	118	64**
8	5	NotA|A|NotA|A|NotA	104	137	244	262	17	104	54**

Note. For those participants for whom the four bound model fits better, typical Group A members fall between the B1 and B2 decision bounds whereas atypical Group A members fall between the B3 and B4 bounds. The Error column shows the value of the error parameter. The –lnL is an indicator of model fit. *χ^2 (2) is significant at $p < .05$; **$p < .01$.

distinct subcategories of Group A members in the A/Not A learning phase. Importantly, this finding suggests that subcategories of a larger group *can* be recognized in the absence of explicit demands to do so. Although the stimuli were generated such that the optimal response was to recognize the two subcategories, two participants ignored the atypical Group A members during the A/Not A learning phase and responded as if only typical Group A members were presented.

Category Responses in the B/Not B Learning Phase. Participants had to learn about Group B before continuing to the subsequent test phase where they make Group A versus Group B categorization judgments. The B/Not B learning phase responses were analyzed with the two bound model that corresponds with a single cluster of "Group B" responses between the two bounds. No comparative model fits were done, as no other models were expected to fit the unimodal, normally distributed Group B data. The optimal models given the stimuli and the two bound models that

TABLE 3. Two bound models that best fit the last three blocks of the presented stimuli (PS) and each participant's responses in the B/Not B learning phase of Study 1

Part #	Model	Block	B1	B2	Error	–lnL
PS	optimal bounds		266	314	16	74
1	NotB|B|NotB	3	253	319	16	61
1	NotB|B|NotB	4	242	319	15	50
1	NotB|B|NotB	5	263	328	20	77
3	NotB|B|NotB	3	271	347	15	54
3	NotB|B|NotB	4	263	342	22	71
3	NotB|B|NotB	5	264	348	28	84
4	NotB|B|NotB	3	267	328	10	38
4	NotB|B|NotB	4	259	334	9	34
4	NotB|B|NotB	5	266	325	10	27
5	NotB|B|NotB	3	257	338	14	40
5	NotB|B|NotB	4	261	332	12	36
5	NotB|B|NotB	5	250	335	24	70
6	NotB|B|NotB	3	284	329	12	61
6	NotB|B|NotB	4	274	333	18	74
6	NotB|B|NotB	5	264	332	24	86
7	NotB|B|NotB	3	255	336	21	63
7	NotB|B|NotB	4	249	339	11	23
7	NotB|B|NotB	5	248	341	33	90
8	NotB|B|NotB	3	259	348	47	115
8	NotB|B|NotB	4	268	330	27	96
8	NotB|B|NotB	5	258	333	35	103

Note. Group B members fall between the B1 and B2 bounds. The –lnL value is an indicator of model fit.

best fit the participants' B/Not B categorization responses are shown in Table 3. Participants had no difficulty in learning about the Group B members. This finding is pertinent because the Group A versus Group B categorization predictions for the A/B test phase are based on reasonably accurate learning about Group B as well as about Group A.

Decision Bound Analysis of the A/B Test Phase. As an indicator of whether or not the atypical Group A members were included in subsequent judgments about Group A, participants' category responses in the A/B test phase were analyzed to determine the decision bound model that best fit their responses. A number of possible models were compared.

Exactly what models make sense for the A/B test phase? First, it would be highly inaccurate for participants to categorize stimuli at the low end of the Group A range as belonging to Group B (whose members had high scholastic abilities). Thus, it was expected that the model that best fit participants' responses in the A/B test phase would have Group A responses to the stimuli with very low scores.

The simplest model tested was a single bound model with Group A corresponding to the low values of scholastic ability and Group B corresponding to the high values. This single bound model is the optimal model regardless of whether participants include or exclude the atypical Group A members as belonging to Group A (see Figure 2, 3rd panel). If they include the atypical Group A members, they should position their single decision bound to include Group A responses to stimuli in the region where atypical Group A scores were presented ($x > 241$). Alternately, if the atypical Group A members are excluded, the position of the bound should be somewhere in the intermediate region between the typical and atypical Group A members ($174 < x < 241$).

In addition to the single bound model, however, two other models were fit to each participant's responses. Both of these models reflect the idea that the atypical Group A members might be contrasted away from the typical Group A members in the test phase. Contrast effects are usually described as an intercategory phenomenon (Corneille, Klein, Lambert, & Judd, 2002; Krueger, Rothbart, & Sriram, 1989; Tajfel & Wilkes, 1963). Nonetheless, there are two reasons why it is worth investigating contrast between the typical and atypical members of Group A. First, if the atypical Group A members were mentally recategorized, then contrast between the typical and atypical Group A members might be expected as an intercategory phenomenon. Second, to the authors' knowledge, no work has been published on contrast effects between the various portions of a disjunctive category. It is possible that a contrast effect would occur between the typical and atypical Group A members even if both were included in the mental representation of the overall category. Such a finding would be interesting from a categorization standpoint and would expand current knowledge concerning the conditions that produce contrast effects.

To test for a contrast effect between the typical and atypical Group A members, two additional models were fit to the A/B test phase data. First, a two bound model with Group A at the low and high ends and Group B in the middle was tested. That is, within the region where atypical Group A and Group B members were presented, the atypical Group A members would be shifted up to the high end of the Group B range so that the model that best fit the responses would be a two bound model (A | B | A). Second, a less extreme contrast model with three bounds was tested. In this model also, the lowest responses would be Group A responses. Within the region where atypical Group A and Group B members were presented, however, the atypical Group A members would be shifted up to the middle of the Group B range such that a three bound model would best fit the responses (A | B | A | B).

The single bound (A | B), two bound (A | B | A), and the three bound (A | B | A | B) models were thus fit to participants' responses in the A/B test phase. The models that best fit each of the three blocks of test data are shown in Table 4. (The three bound model had to be significantly better than the two bound model to be reported as the best fitting model. However, the χ^2 values reported always reflect comparison with the single bound model.)

Of the five participants who recognized two subcategories of Group A in the learning phase (Participants 1, 3, 4, 5, and 8), three participants' responses (Participants 3, 4, 5) were best fit with a model that indicated the atypical Group A members were contrasted away from the typical Group A members. Participant 4 even had two blocks of responses where the atypical Group A members were contrasted all the way to the top of the Group B region such that the two bound A/B/A model fit

TABLE 4. Models that best fit participants' responses in the A/B test phase of Study 1

<table>
<tr><th>Part #</th><th>Block</th><th>Model</th><th>B1</th><th>B2</th><th>B3</th><th>Error</th><th>–lnL</th><th>χ^2</th><th>2 Group A Subcategories?</th></tr>
<tr><td>1</td><td>1</td><td>A|B|A|B</td><td>115</td><td>117</td><td>261</td><td>22</td><td>54</td><td>10**</td><td>Y</td></tr>
<tr><td>1</td><td>2</td><td>A|B|A|B</td><td>81</td><td>97</td><td>268</td><td>15</td><td>48</td><td>62**</td><td>Y</td></tr>
<tr><td>1</td><td>3</td><td>A|B|A|B</td><td>100</td><td>102</td><td>265</td><td>17</td><td>49</td><td>58**</td><td>Y</td></tr>
<tr><td>3</td><td>1</td><td>A|B|A|B</td><td>161</td><td>279</td><td>300</td><td>14</td><td>70</td><td>41**</td><td>Y</td></tr>
<tr><td>3</td><td>2</td><td>A|B|A|B</td><td>160</td><td>262</td><td>297</td><td>18</td><td>88</td><td>13**</td><td>Y</td></tr>
<tr><td>3</td><td>3</td><td>A|B|A|B</td><td>173</td><td>259</td><td>291</td><td>19</td><td>81</td><td>8**</td><td>Y</td></tr>
<tr><td>4</td><td>1</td><td>A|B|A|B</td><td>164</td><td>290</td><td>309</td><td>12</td><td>55</td><td>63**</td><td>Y</td></tr>
<tr><td>4</td><td>2</td><td>A|B|A</td><td>188</td><td>313</td><td>—</td><td>29</td><td>54</td><td>49**</td><td>Y</td></tr>
<tr><td>4</td><td>3</td><td>A|B|A</td><td>195</td><td>310</td><td>—</td><td>26</td><td>52</td><td>57**</td><td>Y</td></tr>
<tr><td>5</td><td>1</td><td>A|B|A|B</td><td>164</td><td>286</td><td>312</td><td>9</td><td>48</td><td>98**</td><td>Y</td></tr>
<tr><td>5</td><td>2</td><td>A|B|A|B</td><td>152</td><td>277</td><td>305</td><td>7</td><td>46</td><td>107**</td><td>Y</td></tr>
<tr><td>5</td><td>3</td><td>A|B|A|B</td><td>170</td><td>262</td><td>299</td><td>23</td><td>90</td><td>7*</td><td>Y</td></tr>
<tr><td>6</td><td>1</td><td>A/B/A</td><td>183</td><td>380</td><td>—</td><td>36</td><td>23</td><td>4*</td><td>N</td></tr>
<tr><td>6</td><td>2</td><td>A/B</td><td>190</td><td>—</td><td>—</td><td>48</td><td>35</td><td>0, ns</td><td>N</td></tr>
<tr><td>6</td><td>3</td><td>A/B/A</td><td>166</td><td>367</td><td>—</td><td>42</td><td>46</td><td>11**</td><td>N</td></tr>
<tr><td>7</td><td>1</td><td>A/B</td><td>153</td><td>—</td><td>—</td><td>9</td><td>5</td><td>0, ns</td><td>N</td></tr>
<tr><td>7</td><td>2</td><td>A/B/A/B</td><td>199</td><td>272</td><td>272</td><td>0.2</td><td>2</td><td>9**</td><td>N</td></tr>
<tr><td>7</td><td>3</td><td>A/B</td><td>168</td><td>—</td><td>—</td><td>14</td><td>4</td><td>0, ns</td><td>N</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>—</td><td></td><td></td><td></td><td></td></tr>
<tr><td>8</td><td>1</td><td>A|B</td><td>0</td><td>—</td><td>—</td><td>702</td><td>118</td><td>2, ns</td><td>Y</td></tr>
<tr><td>8</td><td>2</td><td>A|B</td><td>23</td><td>—</td><td>—</td><td>662</td><td>112</td><td>2, ns</td><td>Y</td></tr>
<tr><td>8</td><td>3</td><td>A|B</td><td>0</td><td>—</td><td>—</td><td>505</td><td>112</td><td>1, ns</td><td>Y</td></tr>
</table>

Note. The –lnL value is a measure of model fit. *χ^2 value comparing that model to the single bound model was significant at the $p < .05$ level; **$p < .01$ level. When neither the two bound nor the three bound model fit better than the single bound model, the $\chi^2(2)$ value comparing the single and three bound models is reported.

best. For these three participants, the atypical Group A members were clearly not ignored when making Group A versus Group B categorization judgments.

Participant 1's responses were best fit by a three bound A/B/A/B model. However, closer inspection of the data for this participant indicates that the responses were more in line with an A/B model, except for a few miscategorized Group B members. These erroneous Group B responses fell in the middle of region of the *typical* Group A members where no Group B scores were ever presented and these responses make the A/B/A/B model fit better. Notice also that Participant 1's upper bounds for Group A members all include the mean ($\mu = 260$) of the atypical Group A members, indicating that this participant included the atypical Group A members in his or her A/B judgments. Overall, it appears that Participant 1 included the atypical group members in Group A categorization judgments and did not contrast the atypical group members away from the typical group members.

Although Participant 8 learned the two subcategories of Group A and the single cluster of Group B members in the learning phases, this participant apparently did not remember the category structures the next day. The model that best fits this participant's responses indicates that Group B responses were given throughout the stimulus region.

In summary, the model fits suggest that four of the five participants who recognized two subcategories of Group A members in the learning phase also included the atypical Group A members as part of Group A when making Group A versus Group B judgments in the A/B test phase. Interestingly, three of them contrasted the atypical Group A members away from the typical Group A members, even though this shift involved a movement *towards* Group B. The fifth participant (Participant 8) classified almost every stimulus as a Group B member, indicating a general state of confusion during the A/B test phase.

For the two participants (Participants 6 and 7) who ignored the atypical Group A members in the A/Not A learning phase, the best fitting model is the single bound (A/B) model in three blocks, and the atypical Group A members were clearly excluded from the Group A categorizations in these cases.

In two blocks (Participant 6, blocks 1 and 3), the model that best fits the responses is a two bound model (A/B/A). However, inspection of these bounds indicates that the value of the decision bound separating the higher Group A responses from the Group B responses (B2) is higher than the highest stimulus presented. These high B2 bounds allow the A | B | A model to accommodate a couple of high but scattered Group A responses. Ignoring these anomalously high B2 bounds, the upper bounds for the Group A responses in these 2 blocks indicate that Participant 6 excluded the atypical Group A members when making Group A categorizations.

Finally, the model that best fits Participant 7's responses in the second block of the A/B test phase is a three bound A | B | A | B model. Once again, however, the three bound model fits better only because of a single Group A response in the region of the Group B stimuli (at $x = 272$). The value of the decision bound that encompasses all other Group A members ($B1 = 199$) clearly excludes the atypical group members.

Not surprisingly, the categorization data suggest that participants who ignored the atypical Group A members in the A/Not A learning phase also ignored the atypical Group A members in a subsequent categorization task.

Decision Bound Analysis of the A/Not A Test Phase. The A/Not A test phase was included in order to see what happened to perceptions of the atypical Group A members in the event that these individuals were recognized but subtyped and later ignored in assessments of Group A. All indications are that, if the atypical Group A members were recognized as a coherent subcategory initially, they were, in fact, included in assessments of the overall group. Thus, the A/Not A test phase becomes irrelevant to the question it was intended to answer. The results of the model fits are thus not discussed except to mention that they were consistent with the results of the A/B test phase.

Estimates of Central Tendency. The analysis next turns to more traditional measures of subtyping. One measure of the inclusion or exclusion of the atypical group members comes from participants' assessments of the "average" Group A member. Recall that the assessments of central tendency and range were assessed on the second day, after both the A/Not A test phase and after the A/B test phase. These two estimates are clearly not independent. However, because the immediately preceding task may affect the estimates, both the estimates following the A/Not A test phase and the estimates following the A/B test phase were analyzed separately.

The central tendency estimates for all participants are listed in Table 5. When the atypical Group A members are clearly recognized as a distinct subcategory of

TABLE 5. Estimates of central tendency (average) and range (low and high) for participants in the A/Not A learning phase of Study 1

Part #	Average A/ Not A	Low A/ Not A	High A/ Not A	Average A/B	Low A/B	High A/B	2 Group A Subcategories?
1	206	78	290	210	94	270	Y
3	186	58	250	142	62	254	Y
4	194	70	326	190	74	314	Y
5	194	74	290	194	90	274	Y
6	102	38	162	86	34	138	N
7	78	38	102	82	118	126	N
8	150	42	366	162	54	366	Y

Group A, are they included or excluded when assessing the overall group? Recall that Participants 1, 3, 4, 5, and 8 recognized two subcategories of Group A in the A/NotA learning phase. For these five participants, a t–test comparing the mean of the estimates of central tendency to the known mean of all of the Group A members ($\mu = 170$) showed that the estimated central tendency did not differ significantly from this known mean, $M = 186$, $t(4) = 1.67$, *ns* following the A/Not A test phase; $M = 180$, $t(4) = 0.79$, *ns* following the A/B test phase.

A t–test comparing the mean of these participants' central tendency estimates to the known mean of the *typical* Group A members ($\mu = 125$) was also conducted. Using a Bonferroni correction ($\alpha = .025$), this test indicated that the mean of the estimated central tendency values was greater than the known mean of the *typical* group members, $M = 186$, $t(4) = 6.38$, $p < .01$ following the A/Not A test phase; $M = 180$, $t(4) = 4.49$, $p < .02$ following the A/B test phase.

Together, these comparisons indicate that these five participants' estimates of central tendency for Group A were influenced by the atypical group members. More generally, this result shows that when participants recognized a distinct subcategory of atypical group members, they did not ignore this atypical subcategory when assessing the overall group. Although the present study is different from traditional subtyping studies in many ways, this study adds to the growing number of failed attempts to provide unequivocal support for the idea that subtyping, and not simple discounting, is responsible for the lack of impact that highly atypical group members have on people's group assessments.

The two participants who ignored the atypical Group A members during the A/Not A learning phase (Participants 6 and 7) also ignored the atypical group members in their assessments of central tendency. In fact, their estimates of Group A's central tendency were lower than the known mean of the typical Group A members ($\mu = 125$) indicating a contrast effect between Group A and Group B. Following the A/B test phase, this contrast effect was shown to be significant by comparing participants' estimates of central tendency (Participant 6's *estimated* $M = 86$, Participant 7's *estimated* $M = 82$) with the known central tendency for the typical Group A members, $\mu = 125$, $t(1) = 20.5$, $p < .05$. For the estimates taken following the A/Not A test phase, there was a marginally significant shift in the Group A estimates away from the mean of Group B, Participant 6's estimated $M = 102$, Participant 7's estimated $M = 78$, $t(1) = 12.0$, $p < 0.1$.

Estimates of Range. The estimates of the low and high values for Group A for those participants who recognized two subcategories of Group A are listed along with the

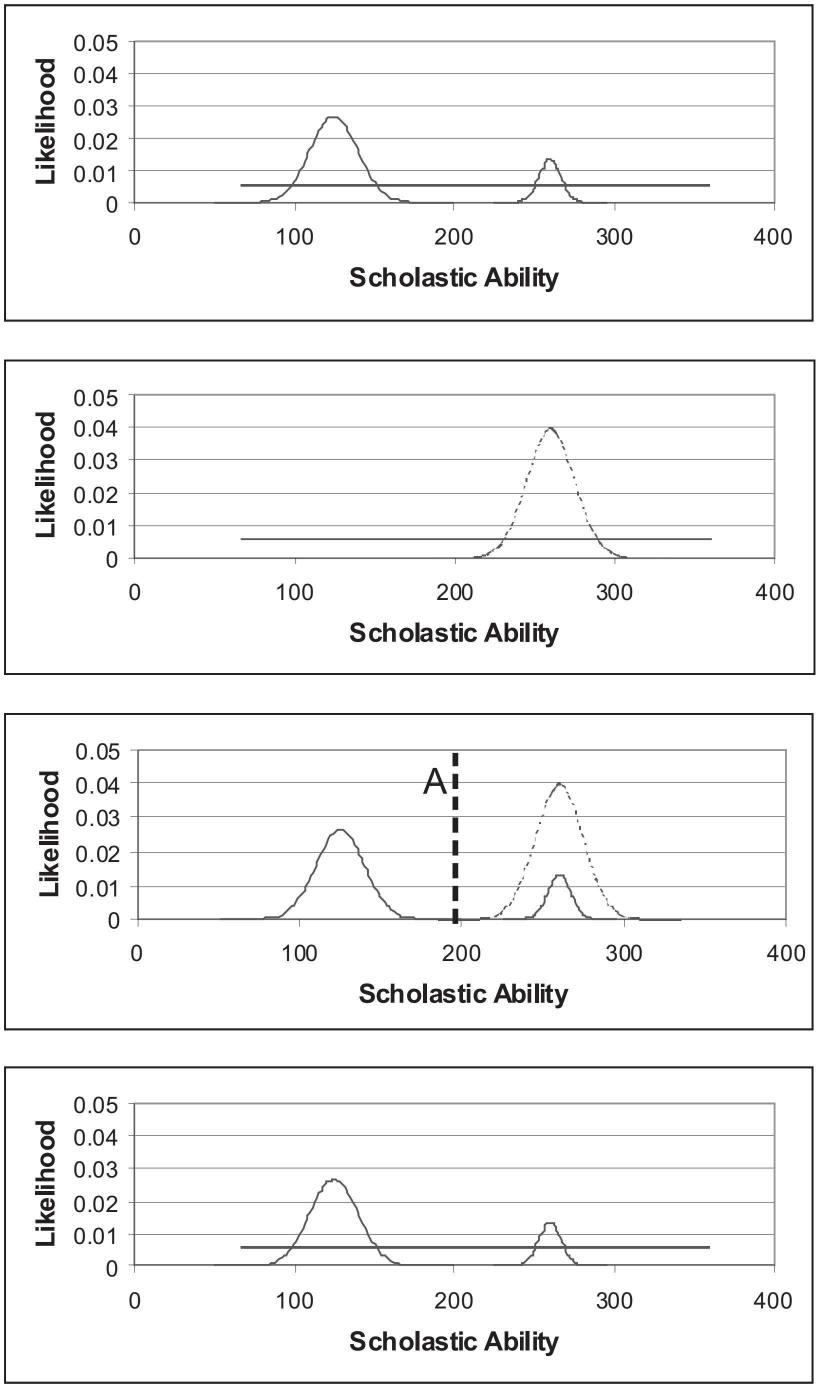

FIGURE 3. Probability density functions for Study 2. Solid function = Group A pdf; Dotted function = Group B pdf; Dashed function = Not A pdf or Not B pdf. The dotted vertical line in the third panel indicates the optimal bound. The solid vertical line indicates the subtyping bound.

estimates of central tendency in Table 5. Of interest is the estimate for the highest Group A member, as this number should give some insight into whether or not the atypical Group A members were included in assessing the overall group.

Participants' estimates of the highest scholastic ability score for Group A can be compared to the lowest actual score for the atypical Group A stimuli. As evidenced in Figure 2 (top panel), there is a large separation between the typical Group A members (highest typical stimulus presented at x = 174) and the atypical Group A members (lowest atypical stimulus presented at x = 241). Thus, the inclusion of even the lowest scores from the atypical Group A members in participants' perceived ranges should be indicative of the atypical Group A members influencing assessments of the overall group.

As the range estimates are not expected to be normally distributed, a signed rank test was used to compare participants' estimates of the high scholastic ability score for Group A to the lowest scholastic ability score actually presented for an atypical Group A member. For the 5 participants who recognized two subcategories of Group A in the A/NotA learning phase, estimates of the highest Group A score were significantly higher than the known lowest score presented for an atypical Group A member, $\Sigma R_{neg} = 0$, $p < .05$, one–tailed. (This test was significant regardless of whether the estimates following the A/Not A test phase were used or those following the A/B test phase were used.) The estimated high score thus indicates that the atypical Group A members were included in assessments of the overall group when two subcategories of Group A were recognized in the A/NotA learning phase.

A signed rank test cannot be performed on the high estimates for the two participants who failed to recognize the atypical Group A members during the A/Not A learning phase because only two participants produced data of this type. However, note that the high estimates for these two participants (Participants 6, *estimated high* = 138; Participant 7, estimated high = 126) are well below the lowest atypical Group A score presented (x = 241) and are, in fact, lower than the highest typical Group A score presented (x = 174). Clearly, the atypical Group A members were ignored by these two participants.

DISCUSSION

The majority of participants in the A/Not A learning phase recognized there were distinct subcategories of typical and atypical Group A members. These participants included the atypical Group A members as part of Group A in the context of a Group A versus Group B categorization task and also included the atypical subcategory in estimates of central tendency and range for the overall group. Thus, when participants clearly recognized the atypical Group A members as a distinct, coherent category, the assessments of the overall group were consistently influenced by the atypical group members. These findings run counter to the idea of subtyping as a means of promoting stereotype maintenance. Instead, the finding of contrast between typical and atypical members of Group A is consistent with the idea that subgrouping (Maurer et al., 1995; Park, Ryan, & Judd, 1992) may promote stereotype change by increasing perceptions of variability.

Two participants failed to recognize the atypical Group A members as a distinct subcategory during the A/Not A learning phase and failed to recognize their existence in any of the subsequent assessments of Group A. They ignored the atypical group members despite (1) constant feedback regarding the existence of the atypical Group A members, (2) a likelihood of occurrence for the atypical Group A members that was higher than the likelihood of occurrence for individuals who were not members of Group A, and (3) a reasonably large ratio of atypical to typical group members (30:60) such that, in terms of frequency of occurrence, the atypical Group A members were not wildly atypical. As there is no evidence that these participants viewed the atypical group members as belonging to any category, the most that can be claimed is that these two participants discounted the atypical Group A members.

Those participants who ignored the atypical Group A members in the A/Not A learning phase apparently also excluded the atypical members in subsequent assessments of Group A. Estimates of central tendency and high value both indicated these participants excluded the atypical Group A members in assessments of the overall group. In addition, these participants did not categorize the atypical Group A members as belonging to Group A in the A/B test phase. This finding is important, because it was possible that participants perceived the likelihood of encountering individuals who were not members of Group A as being higher than the likelihood of encountering atypical Group A members in the A/Not A learning phase. However, they might still have realized there were some atypical Group A members and this knowledge might have displayed itself in the judgments of Group A versus Group B. This was apparently not the case, and results indicated no impact of the atypical Group A members on any of the judgments of Group A, including the A/B test phase categorization judgments. Once discounted, atypical group members are apparently thoroughly discounted.

Given that the majority of the participants in Study 1 included the atypical Group A members in their assessments of Group A on all three measures (categorization, estimate of central tendency, and estimate of high value), it seems reasonable to conclude that creating a subcategory of atypical group members did not lead to the exclusion of the atypical members when assessing the overall group. Of course, because the optimal response according to Decision Bound Theory was to include the atypical Group A members in assessments of Group A, it may not be too surprising that the majority of the participants included the atypical group members when assessing Group A in Study 1.

Study 1 suggests that something more than just the simple recognition of a subcategory of highly atypical group members is required for subtyping to occur. Study 2 investigates the possibility that increased overlap between the characteristics of the atypical subcategory and the characteristics of members of a contrast category may lead to subtyping. More specifically, if the relative likelihoods are set up so that the atypical group members should be ignored in order to produce optimal categorization responses, then subtyping might occur.

STUDY 2

As described earlier, Decision Bound Theory claims that categorization responses are made in a deterministic fashion and are based on the relative likelihood of two categories. If the likelihood of encountering a Group B member with a scholastic ability score of 260 is higher than the likelihood of encountering a Group A member with the same score, then the optimal categorization response when presented with a stimulus score of 260 is to respond "Group B." Thus, Decision Bound Theory suggests the following set of conditions under which a previously learned subcategory of atypical Group A members might later be ignored, or subtyped (Figure 3). Participants would first learn about both the typical and the atypical Group A members during an A/Not A learning phase identical to that of Study 1 (Figure 3, top panel). The subsequent B/Not B learning phase would be similar to that of Study 1, but the mean of the 90 Group B members would be shifted down to coincide with the mean of the 30 atypical Group A members ($\mu = 260$; Figure 3, second panel). Consequently, in the A/B categorization task, the optimal response to scores in the region of $x = 260$ would be to respond "Group B" since the Group B likelihood is higher than the Group A likelihood in this region (Figure 3, third panel). Thus, although the atypical Group A members would originally be recognized as belonging to Group A, the optimal response in a later assessment of Group A (the A/B categorization task) would ignore the atypical group members.

In Study 2 the optimal bound is coincident with the subtyping bound (Figure 3, third panel). That is, if participants perform optimally they should ignore the atypical Group A members in their A/B test phase categorization responses. If they also ignore the atypical group members in their assessments of central tendency and dispersion, subtyping will be demonstrated. An alternate possibility is that perceivers are both optimal in their categorization responses and rational. That is, they may ignore the atypical group members in the context of a Group A versus Group B categorization task, but still acknowledge the existence of the atypical group members when assessing Group A's central tendency and dispersion.

Study 2, then, tests the idea that participants might ignore a previously learned subcategory of atypical group members when subsequent categorization performance is optimized by doing so. In more social terms, atypical group members might be subtyped when their characteristics coincide with the characteristics of a salient contrast group.

METHOD

Participants

Five participants (all female) were recruited from the University of California, Santa Barbara (UCSB) student population and were paid $18 for their participation in the experiment.

Stimuli and Procedure

The stimuli were identical to those used in Study 1, except that the mean of Group B was shifted down to equal the mean of the atypical Group A members ($\mu_{atypicalA} = \mu_B = 260$, see Figure 3). The procedure for Study 2 was identical to that used in Study 1, except that the A/B test phase always preceded the A/Not A test phase. As discussed earlier, the optimal bound in the A/B test phase of Study 2 is also the subtyping bound, and the optimal categorization response is to *not* include the atypical Group A members as belonging to Group A.

RESULTS

Category Responses in the A/Not A Learning Phase. The two bound NotA | A | NotA model and the four bound NotA | A | NotA | A | NotA model were compared in the A/Not A learning phase. For four of the five participants, the four bound model fit

TABLE 6. Best fitting bounds for presented stimuli (PS) and participants' data in the last three blocks of the A/Not A learning phase of Study 2

Part #	Block	Model	B1	B2	B3	B4	Error	–lnL	$\chi^2(2)$
PS		Optimal bounds	106	144	248	272	13	95.0	
1	3	NotA/A/NotA/A/NotA	75	158	238	279	21	100.7	73**
1	4	NotA/A/NotA/A/NotA	80	150	245	297	14	76.2	146**
1	5	NotA/A/NotA/A/NotA	76	140	238	291	13	72.7	166**
2	4	NotA/A/NotA	0	160	—	—	40	67.4	0, ns
2	5	NotA/A/NotA	59	124	—	—	46	80.3	0, ns
2	6	NotA/A/NotA	93	108	—	—	47	39.5	0, ns
3	3	NotA/A/NotA/A/NotA	92	159	252	278	17	96.4	81**
3	4	NotA/A/NotA/A/NotA	93	150	247	273	14	90.5	100**
3	5	NotA/A/NotA/A/NotA	93	150	249	266	12	73.4	110**
4	3	NotA/A/NotA/A/NotA	86	142	254	288	13	81.1	131**
4	4	NotA/A/NotA/A/NotA	92	145	244	283	11	68.6	170**
4	5	NotA/A/NotA/A/NotA	89	146	245	276	9	57.7	178**
5	3	NotA/A/NotA/A/NotA	106	152	241	289	24	132.3	49**
5	4	NotA/A/NotA/A/NotA	106	145	241	279	31	136.8	18**
5	5	NotA/A/NotA/A/NotA	94	141	248	278	38	135.0	9*

Note. Model indicates the best fitting model. In the 2 bound model, Group A responses occurred between B1 and B2. In the 4 bound model, typical Group A responses occurred between the B1 and B2 bounds whereas atypical Group A responses occurred between the B3 and B4 bounds. The –lnL is an indicator of model fit. *$\chi^2(2)$ value comparing the two bound and four bound models is significant at $p < .05$; **Indicates $p < .01$.

significantly better than the two bound model, suggesting that these participants learned two distinct subcategories of Group A members (Table 6, Participants 1, 3, 4, 5). Note that, for these four participants, the region assigned to the atypical Group A members (between B3 and B4) encompasses the mean of the atypical group members ($\mu = 260$) in every case. For one of the five participants, a two bound model fit better (Table 8, Participant 2). This participant categorized the typical Group A members between the two bounds (B1 and B2) and ignored the atypical Group A members in her category responses.

The optimal bounds given the stimuli in the A/Not A learning phase are also shown in Table 6 for comparison. As in Study 1, although four participants' responses suggested that they recognized two subcategories of Group A, the exact positioning of their bounds only approached optimality.

Category Responses in the B/Not B Learning Phase. The B/Not B learning phase data were again analyzed only with the two bound model that corresponds to a single cluster of Group B responses between the two bounds (B1 and B2). The optimal model given the stimuli and the two bound models that best fit the participants' responses are shown in Table 7.

Decision Bound Analysis of the A/B Test Phase. The same three decision bound models compared in the A/B test phase of Study 1 were compared here. The optimal model for these data is a single bound A|B model. An optimal responder would respond "Group A" throughout the region where typical Group A members

TABLE 7. Two bound models that best fit the last three blocks of the presented stimuli (PS) and each participant's responses in the B/Not B learning phase of Study 2

Part #	Block	B1	B2	Error	–lnL
PS		236	284	16	73.5
1	3	233	300	9	31.0
1	4	229	308	14	43.1
1	5	229	305	12	31.9
2	3	238	309	19	68.9
2	4	238	311	8	36.6
2	5	240	306	11	44.0
3	3	206	299	74	142.1
3	4	215	344	120	147.8
3	5	224	316	109	134.4
4	3	217	292	17	54.6
4	4	218	293	13	39.5
4	5	213	299	20	55.2
5	3	231	294	36	112.8
5	4	240	295	36	110.9
5	5	223	292	43	121.1

Note. The –lnL is an indicator of model fit.

were presented and would respond "Group B" throughout the region where scores for both Group B members and atypical Group A members were presented (around $\mu = 260$). In the event that participants did not perform optimally and failed to exclude the atypical Group A members from their Group A categorization responses, the two bound A | B | A model and the four bound A | B | A | B model were also fit to the data. Note that, of these two models, the A | B | A | B model better reflects the structure of the stimuli, because the atypical Group A members were positioned in the center of the Group B distribution. A better fit for the A | B | A model would suggest the atypical Group A members were not only recognized as belonging to Group A, but were also contrasted away from the typical Group A members.

In summary, the single bound (A | B), two bound (A | B | A), and the three bound (A | B | A | B) models were fit to participants' responses in the A/B test phase. The models that best fit each of the three blocks of test data for each participant are shown in Table 8. (As in Study 1, χ^2 values reported are those compared to the single bound model. However, the 3 bound model had to be significantly better than the 2 bound model as well to be considered the best fitting model.)

The four participants who recognized two subcategories of Group A in the A/NotA learning phase (Participants 1, 3, 4, and 5) included the atypical Group A members when making A/B category judgments. Participants 1 and 3 both responded Group A in the region near the atypical Group A mean of $\mu = 260$ and a three bound model (A/B/A/B) best fit their responses. Participant 1's bounds corresponding to the atypical Group A members are shifted slightly below $\mu = 260$, indicating assimilation of the atypical Group A members toward the typical Group A members.

TABLE 8. Models that best fit participants' responses in the A/B test phase of Study 2

Part #	Block	Model	B1	B2	B3	Error	–lnL	χ^2
1	1	A/B/A/B	163	234	263	16	81.1	8*
1	2	A/B/A/B	164	223	261	15	73.3	18**
1	3	A/B/A/B	170	240	265	13	77.1	13**
2	1	A/B	254	—	—	12	52.7	0, ns
2	2	A/B	265	—	—	10	50.5	0, ns
2	3	A/B	264	—	—	9	47.9	0, ns
3	1	A/B/A/B	156	257	284	12	72.6	50**
3	2	A/B/A/B	149	251	274	11	80.4	33**
3	3	A/B/A/B	217	243	277	10	57.6	34**
4	1	A/B/A	210	272	—	6	24.2	115**
4	2	A/B/A	197	272	—	15	52.8	63**
4	3	A/B/A	180	281	—	23	58.1	43**
5	1	A/B	243	—	—	17	50.9	0, ns
5	2	A/B	219	—	—	38	51.4	0, ns
5	3	A/B	240	—	—	19	48.9	0, ns

Note. The –lnL value is a measure of model fit. *The χ^2 value comparing that model to the single bound model was significant at the $p < .05$ level; **Indicates significance at the $p < .01$ level. When neither the two bound nor the three bound model fit better than the single bound model, the χ^2 (2) value comparing the single and three bound models is reported.

Participant 4's responses were best fit with the A/B/A model indicating that this participant included the atypical Group A members as belonging to Group A, but contrasted them away from the typical Group A members. Finally, Participant 5's data were best fit by a single bound (A/B) model, but the position of the bound indicates that Group A responses were made to scores corresponding to the lower portion of the Group B distribution. As no typical Group A members were presented with scores anywhere near this region, it is reasonable to assume that these Group A responses indicate inclusion of the atypical Group A members in the Group A categorization judgments, and that the atypical Group A members have been assimilated toward the typical Group A members.

Unexpectedly, then, these four participants responded suboptimally, apparently failing to recognize that the likelihood of a Group B member was greater than the likelihood of an atypical Group A member in the region near $x = 260$. Two of them flat out ignored the relative likelihoods and two of them either misperceived the positions of the atypical Group A members relative to the Group B members or reinterpreted the relative positions in an attempt to reveal that they knew the atypical Group A members existed.

Interestingly, Participant 2's responses in the A/B test phase indicate that she included the atypical Group A members when making Group A versus Group B categorization judgments, even though she did not categorize the atypical Group A members as belonging to Group A during the A/Not A learning phase.

Decision Bound Analysis of the A/Not A Test Phase. As in Study 1, it is clear that the atypical Group A members were not ignored by most participants. Consequently, checking the A/NotA test phase data to ensure atypical Group A members were not

TABLE 9. Best fitting models for participants' responses in the A/Not A test phase of Study 2

Part #	Block	Model	B1	B2	B3	B4	Error	–lnL	χ^2
1	1	NotA/A/NotA/A/NotA	74	154	234	274	13	67.8	127**
1	2	NotA/A/NotA/A/NotA	58	166	242	278	21	87.0	63**
1	3	NotA/A/NotA/A/NotA	72	162	243	274	17	79.7	84**
2	1	NotA/A/NotA	0	188	—	—	16	15.4	0, ns
2	2	NotA/A/NotA	0	168	—	—	10	11.1	0, ns
2	3	NotA/A/NotA	0	172	—	—	21	21.2	0, ns
3	1	NotA/A/NotA/A/NotA	86	150	249	274	14	82.9	109**
3	2	NotA/A/NotA/A/NotA	84	145	246	271	13	75.0	119**
3	3	NotA/A/NotA/A/NotA	79	162	254	288	26	117.0	41**
4	1	NotA/A/NotA/A/NotA	64	169	288	333	15	51.2	154**
4	2	NotA/A/NotA/A/NotA	0	166	305	350	11	28.8	211**
4	3	NotA/A/NotA/A/NotA	0	153	298	364	13	36.0	239**
5	1	NotA/A/NotA/A/NotA	84	144	195	259	26	130.6	28**
5	2	NotA/A/NotA/A/NotA	69	153	196	251	28	120.0	15**
5	3	NotA/A/NotA/A/NotA	74	150	182	252	26	114.3	22**

Note. The –lnL value is a measure of model fit. *The $\chi^2(2)$ value comparing that model to the two bound model was significant at the $p < .05$ level; **Indicates significance at the $p < .01$ level.

recategorized as belonging to Group B seems unnecessary. Nonetheless, responses in the A/Not A test phase are described here for two reasons. First, these data should provide additional insight into those participants who assimilated the atypical Group A members toward the typical Group A members. Second, these data might provide some insight into Participant 2's behavior.

The models compared were the four bound NotA | A | NotA | A | NotA model and the two bound NotA | A | NotA model. For the four participants who initially recognized two subcategories of Group A (Participants 1, 3, 4, and 5), the four bound model fit significantly better than the two bound model (see Table 9). Thus, these participants continued to distinguish between the two subcategories of Group A, even in the context of a categorization task where no feedback was provided. As was the case in the A/B test phase, Participants 1 and 3 centered the atypical Group A members close to the known mean of $\mu = 260$. Also consistent with the A/B test phase, Participant 4 contrasted the atypical Group A members away from the typical Group A members whereas Participant 5 assimilated the atypical Group A members toward the typical Group A members. The fact that a four bound model best fits Participant 5's responses indicates that the atypical Group A members were not assimilated to the point where they were no longer distinct from the typical Group A members.

Participant 2 apparently reverted to excluding the atypical Group A members when making A/Not A categorization judgments. The upper bound for the Group A members is only slightly higher than the highest score presented for typical Group A members during the learning phase ($x = 170$), indicating that Participant 2 made Group A categorizations only for typical Group A members during the A/Not A test phase.

TABLE 10. Estimates of central tendency (average) and range (low and high) from Study 2

Part #	Average A/ Not A	Low A/ Not A	High A/ Not A	Average A/B	Low A/B	High A/B
1	162	66	254	162	66	246
2	62	22	134	62	30	162
3	150	70	250	162	74	266
4	98	42	354	114	46	322
5	138	86	234	126	70	234

Estimates of Central Tendency. Table 10 shows the central tendency and range estimates for Study 2. Looking first at the estimates for those participants who recognized two subcategories of Group A members in the A/Not A learning phase (Participants 1, 3, 4, and 5), the mean estimate of central tendency is intermediate between the known mean of all Group A members combined ($\mu = 170$) and the known mean of the typical Group A members alone ($\mu = 125$, $M = 141$ following the A/B test phase, $M = 137$ following the A/Not A test phase). The average estimates for these participants do not differ significantly from either of these known means (all t's < 2.4).

The central tendency estimate from Participant 2, who ignored the atypical Group A members in the A/Not A learning phase, was considerably lower than the known mean of the typical Group A members. This participant apparently did not include the atypical Group A members in her assessments of Group A's central tendency.

Estimates of Range. Looking at the estimated values of the highest Group A member for those participants who recognized two subcategories of Group A (Table 10, Participants 1, 3, 4, and 5) indicates that the estimated high was always higher than the lowest Group B member. (Recall the optimal lower bound for Group B in the B/Not B task was $x = 236$.) That is, the high estimate was not restricted to the range of the typical Group A members and, in fact, was not even restricted to the region where the typical Group A members occurred plus the intermediate region where neither Group A nor Group B stimuli were presented. Instead, the estimates of the high Group A member suggest that participants included at least some of the atypical Group A members in their estimates of the highest Group A score.

In contrast, the estimated high for Participant 2, who ignored the atypical Group A members in the A/Not A learning phase, was lower than the highest score presented for any typical Group A member. Thus, the estimated high score indicates that this participant did not include the atypical Group A members in his or her assessments of Group A.

DISCUSSION

The categorization judgments combined with the central tendency and range estimates tell a consistent story. As in Study 1, these data suggest that, when a subcategory of a group is recognized, it leads to subgrouping and thus the inclusion of atypical members in assessments of the overall group, rather than to subtyping. Study 2 reveals that even when an atypical subcategory is clearly recognized and even when optimal responding requires completely ignoring the subcategory in subsequent classifications of group members, participants still do not subtype! Thus, if subtyping is a means to stereotype maintenance, there must be additional factors beyond atypicality and similarity to a comparison group that drive the subtyping phenomenon.

From a categorization standpoint, the results of Study 2 are interesting because they show that there are conditions under which participants perform sub–optimally. That is, participants responded in a manner that did not optimize the likelihood of a correct response, and thus violated the prediction made based on Decision Bound Theory. When presented with a stimulus in the region of $\mu = 260$, participants responded Group A, even though the likelihood of encountering a Group B member was much higher in this region. Note that this finding is also inconsistent with popular similarity–based exemplar models of categorization such as Nosofsky's Generalized Context Model (Nosofsky, 1986). In this model, the similarity between an atypical Group A member and all of the members of Group B is higher than the similarity between an atypical Group A member and all of the members of Group A. Stimuli with the characteristics of atypical Group A members should be classified as belonging to Group B according to the Generalized Context Model. Thus, we have a finding here that would not have been predicted by either of two influential categorization models.[3]

Note that there are several assumptions implicit in the design of Study 2 that led to the prediction of optimal categorization responses based on relative likelihood. First, participants were assumed to be able to compute the relative likelihood of Group A and Group B accurately even though they had never been trained on direct comparisons of Group A versus Group B. Instead, participants learned the relative likelihood of Group A members compared to nonmembers. Similarly, they learned the relative likelihood of Group B members compared to nonmembers. It is not entirely clear that knowledge of these relative likelihoods can be transferred to a judgment involving the relative likelihood of Group A versus Group B. All of the participants' responses, including Participant 2's anomalous responses, may be indicative of just such a difficulty in transferring previously learned relative likelihood knowledge to a novel set of comparisons. To the authors' knowledge, the assumption that comparisons of relative likelihood can be applied in a transitive manner has not been tested. A consideration of this assumption raises an intriguing proposition: Our assessments of a group may be intimately tied to the specific contrast group that was used as a reference when first learning about the group, and the ability to accurately compare that group to novel contrast groups may be limited.

A second assumption was that people can accurately track base rates across different intergroup comparisons. The specific assumption is that people realize Groups A and B occur equally frequently, even though they've never directly com-

3. If the atypical Group A members are actually seen as a category that is independent of the typical Group A member category, the area under the atypical part of the pdf becomes equal to 1 instead of equal to 1/3. This would effectively heighten the atypical Group A pdf to a point where it was higher than the Group B pdf in the region around $x = 260$. It's tempting to suggest that perceivers were, then, optimally responding when they produced "Group A" responses in the region of $x = 260$. This ignores, however, the fact that 3 times as many Group B members were presented than atypical Group A members. According to both Decision Bound Theory and Generalized Context Theory, this base rate difference should cause participants to respond "Group B" in the region of $x = 260$ even if the atypical group members are seen as a totally separate category.

pared the two. It may be, however, that the relative frequencies of groups that are never directly compared are distorted from reality. Perceivers mistakenly may have thought that the Group A members (or the atypical Group A members) were more numerous relative to the Group B members than they really were. Such a misjudgment of relative frequency could have led to the suboptimal responding to the atypical Group A members in Study 2.

A third assumption was that participants would make categorization responses in line with their relative likelihood estimates even in the absence of feedback. It may be, however, that the absence of feedback releases participants from the motivation for accuracy on each trial. After all, the actual level of accuracy cannot be assessed without feedback. Thus, in the absence of feedback, participants may shift to a strategy of responding in a manner that reflects the distributions of the groups presented, sometimes responding Group A and sometimes responding Group B in regions where members from both groups have been seen in the past. That is, although perceivers respond deterministically in a supervised categorization task, they may switch to a probability matching strategy in the absence of feedback. If this were the case, it would certainly suggest that previously learned decision bounds do not fully characterize a perceiver's category knowledge.

In summary, Study 2 produced somewhat surprising results. In contrast with predictions, participants did not subtype the atypical Group A members, even though they made evaluations of Group A in a context where it was optimal to ignore the atypical Group A members. Like Study 1, Study 2 suggests that, once recognized as an atypical subcategory, atypical group members are not ignored in assessments of the overall group. Both Studies 1 and 2 provide support for the idea of subgrouping but fail to provide any evidence supporting the categorization aspect of subtyping.

GENERAL DISCUSSION

At first blush, the paradigm used in the present work seems at odds with prior research on subtyping. The general approach, however, is the same as that taken in studies that have attempted to isolate the categorization aspect of subtyping by encouraging participants to sort the group members and have then measured the effects of that sorting on perceptions of the overall group (Mauer et al., 1995; Park et al., 1992). Despite this similarity with prior research, the use of categorization models and tasks takes the present work beyond prior work in a variety of ways. First, the paradigm strenuously avoided any explicit experimenter–driven demand to create subcategories within the overall group. However, the stimuli were designed to create pressure to subtype. Because prior work on subtyping suggests the extent of atypicality is the most important factor determining whether subtyping occurs or not (Johnston & Hewstone, 1992; Kunda & Oleson, 1997; see Richards & Hewstone, 2001 for a review), the present studies created pressure to subtype simply by making the atypical group members highly atypical. In the present studies, there was absolutely zero overlap between the distribution of typical members and the distribution of atypical members.

Second, in real world situations, there is significant pressure to *include* the atypical group members in the overall group that must be overcome in order to subtype. That is, either the group member is labeled as affiliated with the overall group (e.g., we've been told he's a lawyer) or the group member exhibits some characteristic that is highly associated with the overall group (e.g., we see he has afro–centric features and dark skin and that leads us to include him in the group "African Americans"). This pressure to see the typical and atypical targets as belonging to the same group is present at some level in all subtyping studies. Typically, participants are initially told that all targets belong to the same group. Situations in the real world often lie between a manipulation that mentions category membership once, and the present case where category membership is declared upon learning about each group member. For instance, when learning about African Americans via face to face contact or via media images, group relevant features are apparent throughout learning. Thus, although critics might argue that the additional pressure to designate each atypical Group A member as belonging to Group A makes it impossible to ignore the atypical group members, many social situations may also make it hard to ignore group membership. Moreover, the additional pressure to label atypicals as belonging to Group A apparently was not overwhelming, as evidenced by the finding that some participants never acknowledged the existence of the atypical Group A members.

The declaration of category membership on each trial gives rise to a third feature of the present work that distinguishes it from prior work. Model fits in the A/NotA categorization task provided a test of whether or not participants did, in fact, recognize two distinct clusters of Group A members. Model fits indicated that the majority of the participants in both studies did recognize separate typical and atypical subcategories within Group A.

This finding set the stage for the critical analysis. When participants recognized a distinct subcategory of atypical Group A members, did they then subtype them and exclude them from their assessment of the overall group? The answer to this question was a resounding "No." In a situation where participants reliably demonstrated spontaneous recognition of a distinct subcategory of atypical group members, they did not subsequently exclude that atypical subcategory when assessing the overall group. This was evidenced by the Group A central tendency and range estimates and by the responses in the Group A versus Group B categorization task. Importantly, Study 2 showed that the atypical Group A members were categorized as belonging to Group A even when both decision bound and exemplar models of categorization suggest that the atypical group members should NOT be categorized as belonging to Group A.

Recall that past studies on subtyping have provided quite strong support for the idea that highly atypical group members have minimal impact on assessments of the overall group. This decreased impact of atypical group members might be explained by decreased retrieval of atypical group members at the time of assessment, as suggested by theories of discounting. The essence of subtyping, however, is that it involves *the recategorization and consequent mental isolation of a subset of similarly atypical group members*. Past research attempting to pin down this categorization aspect of subtyping has been less than definitive. Sorting paradigms provide little insight into whether spontaneous recategorization of highly atypical group members occurs and clustered recall measures are fraught with confounding interpretations that have been dismissed only through failure to reject multiple null hypotheses (and power has not been controlled). Thus, support for the categorization aspect of subtyping is weak.

In attempting to fill this gap, the present research follows the established method of (1) getting participants to recognize a distinct category of atypical group members and then (2) testing the impact of that category of atypical group members on assessments of the overall group. However, the present paradigm avoids explicitly directing participants to create separate categories based on the differences between the typical and atypical group members. In addition, the present research provides a novel test of the categorization aspect of subtyping. Overwhelmingly, the data from Study 1 and Study 2 show that, when participants recognized a subcategory of atypical Group A members, it led to subgrouping (and inclusion in assessment of the overall group) rather than to subtyping (and exclusion when assessing the overall group). This result certainly does not provide support for the categorization aspect of subtyping. Instead, it adds to the growing doubt about whether subcategorization is the mechanism by which stereotypes are maintained despite exposure to highly atypical group members.

In sum, the research reported here attempted to illuminate the categorization aspect of subtyping. Given the results of prior work, it seems safe to say that highly atypical group members are sometimes underrepresented in assessments of the overall group. Yet, the body of evidence supporting the contention that the mechanism behind this effect is categorization–based is not so solidly supported. The present research does not provide support for the idea that recategorization of highly atypical group members leads to their being ignored in assessments of the overall group. Instead, it suggests that if an atypical subcategory is recognized, the outcome will be subgrouping such that assessments of the overall group are impacted by the atypical group members. Obviously, this research does not disprove the possibility that, under some circumstances, recategorization of atypical group members might lead to stereotype maintenance. The challenge is to find those circumstances and show strong evidence of an underlying categorization mechanism.[4] Taking a theory–driven categorization approach to the issue of stereotype change does suggest that evidence of subgrouping is more easily obtained than evidence of subtyping. In the end, the present research raises the challenge to provide irrefutable evidence that subtyping is more than the simple discounting of highly atypical group members.

4. Kunda and Oleson's (1995) studies provide what might be considered the best evidence of perceivers applying the attributes of one highly atypical group member to another highly atypical group member. However, this may simply be reliance on perceived feature covariation (e.g., introverted, small firm lawyer used to predict another lawyer from a small firm is also introverted). Whether one wants to claim this is a new category is an issue open to debate.

REFERENCES

Anderson, J. R. (1980). *Cognitive psychology: Its implications.* San Francisco, CA: Freeman.

Ashby, F. G. (1992a). Multidimensional models of categorization. In F. Gregory Ashby (Ed.), *Multidimensional models of perception and cognition.* Scientific psychology series (pp. 449–483). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ashby, F. G. (1992b). Multivariate probability distributions. In F. Gregory Ashby (Ed.) *Multidimensional models of perception and cognition.* Scientific psychology series (pp. 1–34). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33–53.

Ashby, F., & Maddox, W. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 50–71.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity–choice model. *Psychological Science, 5,* 144–151.

Biernat, M., & Kobrynowicz, D. (1997). Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology, 72,* 544–577.

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., 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.

Darley, J. M., & Gross, P. H. (1983). A hypothesis–confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44,* 20–33.

Devine, P. G., & Baker, S. M. (1991). Measurement of racial stereotype subtyping. *Personality and Social Psychology Bulletin, 17,* 44–50.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*, 134–40.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200.

Gurwitz, S. B., & Dodge, K. A. (1977). Effects of confirmations and disconfirmations on stereotype–based attributions. *Journal of Personality and Social Psychology, 35,* 495–500.

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

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37,* 25–38.

Hewstone, M., Hopkins, N., & Routh, D. A. (1992). Cognitive models of stereotype change: I. Generalization and subtyping in young people's views of the police. *European Journal of Social Psychology, 22,* 219–234.

Hewstone, M., Johnston, L., & Aird, P. (1992). Cognitive models of stereotype change: II. Perceptions of homogeneous and heterogeneous groups. *European Journal of Social Psychology, 22,* 235–249.

Hewstone, M., Macrae, C. N., Griffiths, R., & Milne, A. B. (1994). Cognitive models of stereotype change: 5. Measurement, development, and consequences of subtyping. *Journal of Experimental Social Psychology, 30,* 505–526.

Higgins, E. T., & Lurie, L. (1983). Context, categorization and recall: The "change–of–standard" effect. *Cognitive Science, 15,* 525–547.

Huttenlocher, J., Hedges, L.V., & Vevea, J. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129,* 220–241.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: III. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28,* 360–386.

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.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology, 68,* 565–579.

Kunda, Z., & Oleson, K. C. (1997). When exceptions prove the rule: How extremity of deviance determines the impact of deviant examples on stereotypes. *Journal of Personality and Social Psychology, 72,* 965–979.

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology, 69,* 812–824.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill–defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 128–148.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59,* 173–191.

Park, B., Ryan, C. S., & Judd, C. M. (1992). Role of meaningful subgroups in explaining differences in perceived variability for in–groups and out–groups. *Journal of Personality and Social Psychology, 63,* 553–567.

Posner, M. I. (1974). *Cognition: An introduction.* Glenview, IL: Scott Foresman.

Queller, S., Mackie, D. M., & Stroessner, S. J. (1996). Ameliorating some negative effects of positive mood: Encouraging happy people to perceive intragroup variability. *Journal of Experimental Social Psychology, 32,* 361–386.

Razran, G. (1950). Ethnic dislikes and stereotypes: A laboratory study. *Journal of Abnormal and Social Psychology, 45,* 7–27.

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5,* 52–73.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76,* 45–48.

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,* 81–104.

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 39,* 590–598.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1,* 149–178.

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

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). Hillsdale, NJ: Erlbaum.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45,* 961–977.

Wilder, D. A. (1984). Intergroup contact: The typical member and the exception to the rule. *Journal of Experimental Social Psychology, 20,* 177–194.