Title

Stereotype Relevance Moderates Category Activation: Evidence From the Indirect Category Accessibility Task (ICAT)

Journal

Social Psychological and Personality Science, 1(4)

ISSN

1948-5506

Authors

Stroessner, Steven J
Haines, Elizabeth L
Sherman, Jeffrey W
et al.

Publication Date

2010-10-01

DOI

10.1177/1948550610372144

Peer reviewed

Stereotype Relevance Moderates Category Activation: Evidence From the Indirect Category Accessibility Task (ICAT)

Steven J. Stroessner, Elizabeth L. Haines, Jeffrey W. Sherman and Cara J. Kantrowitz

Social Psychological and Personality Science 2010 1: 335 originally published online 2 August 2010
DOI: 10.1177/1948550610372144

Social Psychological and Personality Science
1(4) 335-343

DOI: 10.1177/1948550610372144
http://spps.sagepub.com

Stereotype Relevance Moderates Category Activation: Evidence From the Indirect Category Accessibility Task (ICAT)

Steven J. Stroessner[1], Elizabeth L. Haines[2], Jeffrey W. Sherman[3], and Cara J. Kantrowitz[1]

Abstract

The impact of behavioral stereotypicality on category accessibility was examined using a novel method, the Indirect Category Accessibility Task (ICAT). In the ICAT, participants learn to distinguish visual stimuli from two categories based on feedback. In two studies, participants were exposed to images of individuals behaving consistently, inconsistently, or irrelevantly with traditional gender stereotypes. ICAT learning was superior in the stereotype-consistent and stereotype-inconsistent conditions compared to the stereotype-irrelevant conditions. These results demonstrate that category relevance moderates category accessibility. Implications for social categorization and stereotype change models are discussed.

Keywords

categorization, behavior stereotypicality, accessibility

Categorization is a fundamental process in intergroup perception (Allport, 1954; Macrae & Bodenhausen, 2000). Once a person is categorized as a member of a social group, stereotypes associated with the group become accessible, influencing attention, elaboration, judgment, and behavior (see Schneider, 2004, for a review). Although the *consequences* of category activation have received ample attention, less is known about what causes categories to become activated in the first place.

Most research on category activation has focused on perceiver and contextual factors. Studies show that categorization is affected by perceivers' available cognitive resources (Gilbert & Hixon, 1991), processing goals (Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997), and motives (e.g., Lepore & Brown, 1997). Categories that are chronically accessible also are activated more readily and effortlessly compared to categories low in accessibility (Blanz, 1999; Bruner, 1957). Categories that are contextually unusual (Taylor, Fiske, Etcoff, & Ruderman, 1978) or in comparison to cultural norms (Stroessner, 1996; Zárate & Smith, 1999) are more likely to be activated than "normative" categories.

Less work has addressed how a *target's characteristics* influence category activation. The present research addresses the effects of one target variable—behavioral stereotypicality—on category accessibility. We expect that whether a target acts typically, neutrally, or atypically in relation to prevailing stereotypes will affect category activation. This is an important issue, as social category activation based on typicality is central in theories of stereotype change. Rothbart and John (1985), for example, argued that targets seen as atypical of their group might not activate their social category. Given prevailing racial stereotypes, a "Black scientist" might not activate his racial category precluding modification of racial stereotypes based on stereotype-disconfirming information. Consistent with this view, Rothbart and Lewis (1988) showed that perceivers were less likely to generalize features exhibited by atypical "frat boys" to their fraternity as a whole compared with stereotypical members.

Stereotype-Consistent Behavior and Category Accessibility

Other theories echo the notion that stereotypical exemplars will most likely activate their correspondent social category. Models of both object (Hintzman, 1986; Medin & Smith, 1984; Posner & Keele, 1968; Rosch & Mervis, 1975) and social categorization (Cantor & Mischel, 1979; Smith & Zárate, 1992) predict that as exemplar typicality increases, the

[1] Columbia University, New York, NY, USA
[2] William Paterson University, Wayne, NJ, USA
[3] University of California, Davis, Davis, CA, USA

Corresponding Author:
Steven J. Stroessner, Columbia University, Barnard College, 3009 Broadway, New York, NY 10027
Email: ss233@columbia.edu

activation of a pertinent category will also increase. In the domain of stereotyping, Rothbart, Sriram, and Davis-Stitt (1996) showed that individuals exhibiting stereotype-consistent traits are judged as better exemplars of their category than individuals displaying stereotype-irrelevant or stereotype-inconsistent information. Locke, Macrae, and Eaton (2005) demonstrated that racial categories are made accessible by typical- rather than atypical-looking category members. Response times to name the ethnicity of common Asian and Caucasian names were faster when faces typical rather than atypical for the category appeared alongside the names. Racial categories were activated in response to typical-looking category members.

Stereotype-Inconsistent Behavior and Category Accessibility

There are theoretical bases for predicting the opposite effect (i.e., greater processing of and greater attention to atypical or counterstereotypical exemplars). Unusual or atypical information can create a subjective feeling of disfluency (Johnston & Hawley, 1994) and draw attention and scrutiny (e.g., Johnston & Hawley, 1994; Tulving, Markowitsch, Kapur, Habib, & Houle, 1994). Norm theory (Kahneman & Miller, 1986) also argues that counterstereotypic category members may be particularly likely to evoke a countervailing category norm, thereby increasing the accessibility of the category. Each of these literatures helps explain why attention tends to be drawn toward information inconsistent with stereotypes (Allen, Sherman, Conrey, & Stroessner, 2009; Sherman, Lee, Bessenoff, & Frost, 1998), although these studies have typically not included measures of category accessibility. Nonetheless, these lines of research suggest that counterstereotypic category members might be particularly likely to increase the accessibility of the category.

Stereotype-Relevant Behavior and Category Accessibility

A third possibility is that both typical and atypical category members induce greater category accessibility than category members who are neither typical nor atypical. Research on schemas and expectancies has shown that information that *either* confirms *or* disconfirms category knowledge is more likely to draw attention than schema- or expectancy-irrelevant information (see Fiske & Taylor, 1984; Srull & Wyer, 1989). This raises the possibility that both highly typical and atypical category members will lead to greater category accessibility than members who are neither.

Challenges in Measuring Social Category Activation

Before detailing how we distinguished these possibilities, we turn to a discussion of existing methods for assessing category accessibility. In most studies examining social categorization and behavior, category labels are provided (with or without participants' awareness) and category accessibility is assessed using trait ratings (for a review, see Kunda & Spencer, 2003). Stereotype use is typically assumed to reflect category activation. Despite this procedure's usefulness, two serious issues prevent its use for studying spontaneous social categorization. First, category activation (accessibility in one's mind) does not always lead to stereotype application (use of the stereotype in judgment). Indeed, stereotypes can be activated without application when there is a motivation to control prejudice (e.g., Devine, 1989) or when external attributions are made for stereotyped behavior (Sherman, Stroessner, Conrey, & Azam, 2005). Second, the use of category labels may prompt categorization processes that might not have occurred without such labels (Macrae & Bodenhausen, 2000).

One method that avoids these problems is the "who said what?" paradigm (Taylor et al., 1978). In this method, participants observe a conversation among members of two groups, and participants later match each comment to the speaker who made it. If matching errors within a category exceed errors between categories, it has been argued that speakers must have been categorized as group members. This paradigm has been used extensively as a measure of spontaneous social categorization, and it has been the primary method for assessing the impact of behavioral information on social categorization (typically, through a manipulation of conversational topic; see Klauer & Wegener, 1998). Although this method has provided useful insights, it is limited in several respects. First, because physical features tend to covary with group membership, matching mistakes might reflect feature-based confusions rather than categorization errors (e.g., Blair, Judd, & Fallman, 2004). Second, the traditional implementation of the method fails to disentangle item, category, and person discrimination in memory (Klauer & Wegener, 1998). Third, the method is not an effective gauge of online variations in category accessibility. Speakers are matched with comments after all stimuli have been presented, so the point at which any category becomes accessible cannot be measured. Although there has been a rich tradition of research using the "who said what?" paradigm, several important issues are difficult to address with this method.

The ICAT

In response to these concerns, we have developed the Indirect Category Accessibility Task (ICAT), a variation of the classic two-alternative forced choice learning paradigm with feedback (e.g., Huber, Shiffrin, Lyle, & Quach, 2002; Monaci, Menegaz, Susstrunk, & Knoblauch, 2004). In the ICAT, participants are presented with a series of pictorial stimuli and told they belong to one of two groups (Groups A and B). After seeing each picture, participants indicate the group to which they believe the stimulus belongs and are provided with feedback for that picture. The participants' task is to learn the rule that distinguishes the stimulus groups.

The logic of this method is as follows: If stimuli increase the accessibility of a category useful for differentiating groups of stimuli, then differences in the ease of learning the differentiating rule reflect at least in part differences in category accessibility. The ICAT (a) avoids using category labels that prompt categorization, (b) allows use of complex, realistic stimuli, (c) permits estimation of when a category becomes accessible, and (d) allows comparison of verbal reports with learning performance.

Overview

The goal of this research was to test the competing theories regarding the relation between behavioral stereotypicality and category accessibility. To do so, we varied the stereotypicality of behavior performed by targets in the ICAT. To the extent that behavioral depictions cause gender categories to become accessible, the more quickly participants should be able to solve the categorization task. Experiment 1 represents a validation study of the ICAT to assess whether an established manipulation that modulates category accessibility—priming—affects performance on this novel learning task. Experiments 2 and 3 then use the ICAT to address the effect of behavioral stereotypicality on category accessibility.

Experiment 1

We began by performing an experiment to validate the ICAT as a measure of gender category accessibility. In one condition, we exposed participants to primes designed to heighten the accessibility of the category ''female.'' Participants in a control condition were exposed to neutral primes. Primes were administered subliminally to eliminate demand characteristics.

Method

Participants. In exchange for $5, 43 undergraduates (8 male, 35 female) participated. Participants were randomly assigned to the female prime condition ($n = 22$) or the control condition ($n = 21$).

Materials and procedures. Participants were told that they would complete two experiments, one on a computer (the priming manipulation) and another involving a puzzle they must solve (the ICAT). Participants were told that the computer task assessed their ability to identify the location of a stimulus flashed on the screen. This task used standard parafoveal priming procedures (Bargh & Pietromonaco, 1982; Macrae, Bodenhausen, & Milne, 1995) to prime either gender or no particular social category (a control condition). In both conditions, participants were asked to focus on a fixation point in the center of a computer screen and to indicate using the keyboard (D = left location, K = right location) where a flash appeared. Each flash was composed of one of the prime stimulus words (presented 75 ms) followed immediately by a consonant mask (75 ms). Each stimulus word appeared in one of four quadrants on the screen, approximately 4 cm from the fixation point. To keep the stimuli within the parafoveal visual field (from 2° to 6° of visual angle), the distance between the participants' eyes and the fixation point was held constant at 57 cm.

In the female prime condition, 9 stimulus primes were presented (SHE, WOMAN, HER, LADY, MISSUS, MISS, GIRL, FEMALE, MA'AM) with a set of 9 filler words (e.g., TRASH, TABLE, LIGHT). In the control condition, stimuli were 18 filler words, the 9 from the female prime condition and 9 more that were similar in length and frequency.

Next, participants began the ICAT. They were told that the task involved ''learning criteria for the categorization of a set of pictures.'' They viewed pictures from two groups, Groups A and B, with half of the pictures belonging to each group. They were to decide whether a picture belonged to Group A or B and to infer the rule that determined group membership. (The actual rule was that Group A pictures contained a girl or woman and Group B pictures did not.) A total of 24 black and white photographs or drawings collected from online databases and illustrated books were used in this task. Half of these pictures contained a representation of at least one female and half did not (i.e., they showed males, animals, or inanimate objects). The pictures containing women showed females engaged in gender-neutral behavior (e.g., walking, reading, eating a meal).

Each picture was displayed by an experimenter for 8 s, and participants indicated the group (A or B) on an answer sheet. After indicating their response, participants were told the correct category by the experimenter and marked their sheet to indicate whether their response had been correct. They were told that they could stop responding when they had provided 10 correct responses in a row (i.e., learned to criterion).

After showing all 24 pictures, the experimenter continued to show the set of pictures two more times (in a different random order), providing a maximum of 72 trials. Participants continued to provide responses until they had 10 in a row correct or until the trials were finished. After the trials, participants were queried regarding the categorization rule.

Before being debriefed, all participants answered several questions to determine whether they could identify the priming stimuli or detect the relation between the two tasks. Although some students expressed suspicion that the flashes contained words, no student was able to accurately name any of the words that were used. In addition, no student expressed the belief that the computer and ICAT tasks were in any way related.

Results and Discussion

Participants' response sheets were coded to assess the degree of rule learning (indicated by 10 correct sequential responses) and the ease of learning (reflected in the ordinal position of the first of the 10 correct responses).[1] Differences in learning across conditions were analyzed by examining the percentage of individuals who learned to criterion and (given that some individuals never learned to criterion) in the mean ranks in trials to learning. The cumulative percentage frequency distribution provided in Figure 1 shows learning over trials, with the percentage of individuals who learned to criterion indicated by the maximum height of each line.

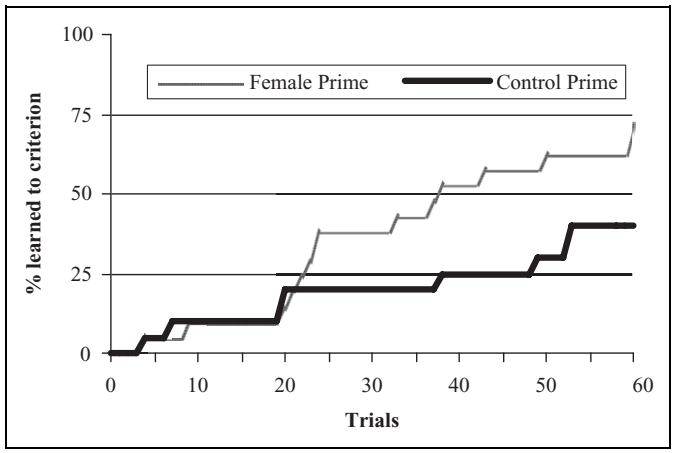

Figure 1. Cumulative percentage frequency distribution of learning over trials on the Indirect Category Accessibility Task as a function of prime condition, Experiment 1

Degree of learning. Because the dependent measure was dichotomous (i.e., the presence or absence of 10 correct sequential responses), logistic regression was used to assess differences in learning between conditions. The degree of learning differed significantly between the two conditions, $\beta = 1.46$, $SE = 0.67$, Wald's $\chi^2(1) = 5.00$, $p < .05$, odds ratio (OR) = 4.29, with better performance in the female prime (68%) than in the control condition (33%).

Ease of learning. The ease of learning was assessed by converting the ordinal position of the first trial in the sequence of 10 correct responses to ranks and analyzing these ranks with a Mann–Whitney U test. This test revealed that learning occurred more quickly in the female prime condition compared with the control condition ($Z = –2.03$, $p < .05$).

Rule reports. Participants' "rules" used to distinguish Group A from Group B pictures were coded by a blind experimenter using both "gist" (i.e., indicating any use of "female" or a synonym) or "strict" (i.e., indicating that the presence of a female indicated Group A membership) criteria. Rule accuracy was low for both "gist" coding (46.7% vs. 37.5% for female prime and control, respectively) and "strict" coding (33.3% vs. 37.5%) and did not differ by condition. This low accuracy, even for participants whose performance indicated that they had "learned" the rule, demonstrates that categorization processes captured by the ICAT might not be readily available for verbal report.

These findings show that the ICAT is useful for indirectly measuring category accessibility. Participants exposed to "female" primes were more likely to learn the gender-based rule that differentiated the stimulus sets. This occurred even though participants were unaware of the primed category and could not easily articulate the critical feature underlying categorization.

Experiment 2

Experiment 2 used the ICAT to test competing hypotheses regarding the effects of behavioral stereotypicality on category accessibility. Participants were presented with images of females engaged in behavior that was stereotypically neutral (as in Experiment 1), stereotypically consistent (e.g., helping a child), or stereotypically inconsistent (e.g., fighting a fire). Stimuli reflecting male behavior were also developed. In both cases, images of males and females were embedded in an array of stimuli not containing representations of males and females, respectively, and participants were asked to learn the rule separating the categories of stimuli based on feedback.

The three hypotheses developed earlier make differing predictions about learning performance across the conditions. If typicality predicts category accessibility, learning a gender-based rule should be best in the stereotype-consistent condition and poorest in the stereotype-inconsistent condition. If category accessibility is driven by novelty, however, learning should be best in the stereotype-inconsistent condition and poorest in the stereotype-consistent condition. If, however, category relevance moderates category accessibility, learning should be better in the stereotype-consistent and stereotype-inconsistent conditions than in the neutral condition.

Method

Participants. A total of 243 volunteer undergraduates from a coed university (males = 93, females = 150) were randomly assigned to one of six conditions created by crossing the stereotypicality (stereotype consistent, neutral, stereotype inconsistent) and target sex (male, female) variables.

Materials. Six pictures of women performing gender-neutral behavior were added to the 14 pictures used in Experiment 1. Two additional sets of 20 pictures each were created to reflect women performing stereotype-consistent (e.g., baking bread) and stereotype-inconsistent behavior (e.g., operating a tow truck). These sets were rated by pretesters ($n = 10$) as varying in stereotypicality ($M = 5.7$, $SD = 0.4$; $M = 4.0$, $SD = 0.3$; $M = 2.3$, $SD = 0.5$, out of 7 for the stereotype-consistent, neutral, and stereotype-inconsistent conditions, respectively). Three sets of 20 pictures each were created showing men engaged in stereotype-consistent, neutral, or stereotype-inconsistent behavior ($M = 5.6$, $SD = 0.3$; $M = 4.0$, $SD = 0.4$; $M = 2.2$, $SD = 0.4$, respectively). The same Group B foils from Experiment 1 were used in the male target conditions, except the pictures of men were replaced with pictures of women engaged in gender-neutral behavior.

Procedures. The procedure was identical to the ICAT phase of Experiment 1 except there were 40 pictorial stimuli (for a maximum of 120 trials).

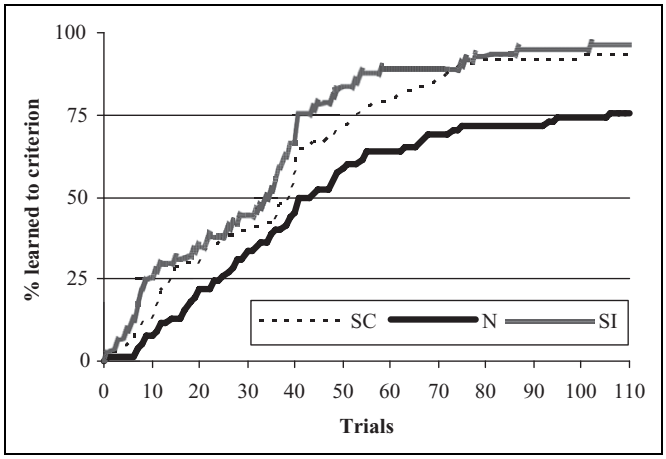

Figure 2. Cumulative percentage frequency distribution of learning over trials as a function of stereotypicality, Experiment 2
Note: SC = stereotype consistent; N = stereotype neutral; SI = stereotype inconsistent.

Results and Discussion

Preliminary analyses. Before performing analyses focused on stereotypicality of behavior, preliminary analyses assessed the impact of participant sex and target sex on both the degree and ease of learning. No main effects or interactions involving participant sex (all $ps > .29$) or target sex (all $ps > .46$) were significant. Similarly, neither variable significantly affected the ease of learning (all $ps > .28$ and .16, respectively).[2] Therefore, these variables were excluded from subsequent analyses.

Degree of learning. Hierarchical logistic regression was used to assess the influence of stereotypicality on degree of learning. Because this was a three-level variable, polynomial dummy codes were created to represent linear and quadratic contrasts for this factor. This approach is particularly appropriate as the linear and quadratic polynomial provide specific tests of the hypotheses based on (a)typicality and stereotype-relevance predictions, respectively.

The overall learning rate was 88.5% and differed significantly by condition, Wald's $\chi^2(2) = 15.74$, $p < .001$. Although the test of the linear pattern was not significant, $\beta = 0.45$, $SE = 0.51$, Wald's $\chi^2(1) = 0.76$, $p = .38$, OR = 1.57, the quadratic component was highly significant, $\beta = 1.51$, $SE = 0.37$, Wald's $\chi^2(1) = 16.53$, $p < .001$, OR = 4.53. As can be seen in Figure 2, learning was higher in the stereotype-consistent and stereotype-inconsistent conditions than in the neutral condition.

Ease of learning. The ordinal positions of the first trial in the sequence of correct responses were submitted to a Kruskal–Wallis H analysis. This yielded a significant stereotypicality effect, $\chi^2(2, N = 243) = 13.80$, $p = .001$, and Mann–Whitney U tests showed that learning occurred earlier in the stereotype-consistent and -inconsistent conditions than in the neutral condition (both $ps < .05$). The comparison between the stereotype-consistent and -inconsistent conditions was marginally significant, $\chi^2(1, N = 165) = 2.91$, $p = .088$, with learning appearing somewhat easier in the stereotype-inconsistent condition.

These results demonstrate that category accessibility varies as a function of behavioral stereotypicality. Learning a gender-based rule was lower and slower when targets engaged in sex-irrelevant behavior than when they engaged in behavior that was *either* typical *or* atypical of gender stereotypes. These findings occurred for male and female participants and for male and female targets. However, the marginally significant difference in the ease of learning between the stereotype-consistent and stereotype-inconsistent conditions indicates that responses to novelty might have especially facilitated category learning.

Table 1. Means and Standard Deviations for Stimulus Sets in Stereotypicality Conditions, Experiment 3

Condition	*M*	*SD*
Extremely stereotype consistent	5.4	0.1
Moderately stereotype consistent	4.7	0.1
Neutral	4.0	0.2
Moderately stereotype inconsistent	3.3	0.2
Extremely stereotype inconsistent	2.6	0.2

Note: 1 = *not at all stereotypical*, 7 = *very stereotypical.*

Experiment 3

Experiment 3 sought to replicate Experiment 2 with improved stimulus sets. Although pilot testing had shown that the sets used in Experiment 2 differed in terms of their consistency with gender stereotypes, even small differences in within-category variability can affect category learning (Fried & Holyoak, 1984; Hahn, Bailey, & Elvin, 2005). In creating these new homogeneous stimulus sets, we also increased the number of sets (and conditions) to five. Using five rather than three levels of stereotypicality allowed testing of patterns that could not be detected with a three-level variable (i.e., cubic or quartic effects).

Method

Participants. A total of 111 undergraduate volunteers (males = 11, females = 100) were randomly assigned to one of the five conditions.

Materials. New stimulus sets were developed by pilot testing 200 pictures of women performing various behaviors. Twelve students judged the degree that each behavior was consistent with traditional gender stereotypes. From these ratings, five sets of 14 pictures were created that were homogeneous in stereotypicality and equally discrepant from the scale midpoint (see Table 1). The foils used in Experiment 2 were utilized with the addition of pictures of two men engaged in gender-neutral behavior from Experiment 2.

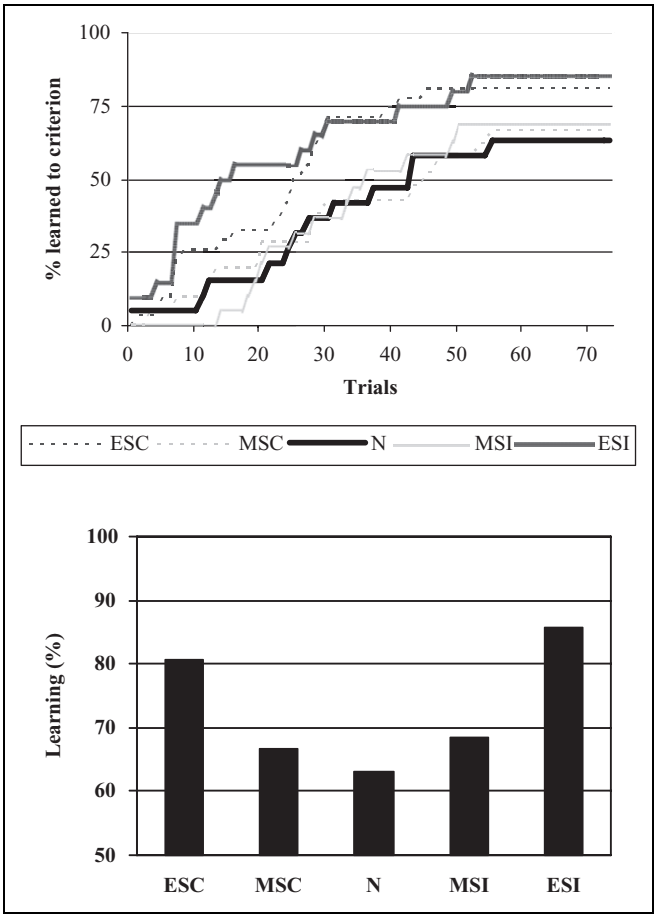

Figure 3. Cumulative percentage frequency distribution of learning over trials (top panel) and degree of learning (bottom panel), Experiment 3
Note: ESC = extremely stereotype consistent; MSC = moderately stereotype consistent; N = stereotype neutral; MSI = moderately stereotype inconsistent; ESI = extremely stereotype inconsistent.

Procedures. The procedures were similar to those of Experiment 2, except two new sets of stimuli were used. Each set contained 28 pictures (14 containing images of women and 14 foils), and the maximum number of trials was 84. Also, after completing all trials, participants were asked to write down the rule that they believed determined whether a picture was in Group A or B and to indicate their confidence in that rule (1 = *not at all sure*, 100 = *absolutely sure*).

Results and Discussion

Degree of learning. Overall, 73.9% of participants learned to criterion, but learning differed by stereotypicality (see Figure 3).[3] Of the four polynomial contrasts tested, only the quadratic component was significant, $\beta = 1.04$, $SE = 0.52$, Wald's $\chi^2(1) = 4.05$, $p < .05$, OR = 2.83, a pattern quite evident in the graph displayed in the lower panel of Figure 3. As in Experiment 2 but with a different set of stimuli, learning was lowest in the neutral condition and highest in the extremely stereotype-consistent and -inconsistent conditions.

Table 2. Accuracy and Confidence Judgments for Participants Who Learned to Criterion, Experiment 3

Condition	Accuracy (%)	Confidence
Extremely stereotype consistent ($n = 25$)	36.0	55.8
Moderately stereotype consistent ($n = 14$)	7.1	38.2
Neutral ($n = 12$)	16.7	42.5
Moderately stereotype inconsistent ($n = 13$)	7.7	26.9
Extremely stereotype inconsistent ($n = 18$)	27.8	60.8

Ease of learning. The Kruskal–Wallis H analysis of trials to learning produced a significant stereotypicality effect, $\chi^2(4, N = 111) = 11.20$, $p < .05$. Mann–Whitney U tests showed that learning was marginally easier in the extremely stereotype-consistent condition compared with the moderately consistent, neutral, and moderately inconsistent conditions (all ps $< .10$) and significantly easier in the stereotype-inconsistent compared with those same three conditions (all ps $< .05$). In contrast to the results of Experiment 2, learning rate was equivalent in the extremely stereotype-consistent and extremely stereotype-inconsistent conditions ($p > .35$).

Rule accuracy and confidence. Although the gist coding of rule accuracy produced no significant differences between conditions, the stricter coding did produce a significant quadratic component, $\beta = 1.39$, $SE = 0.67$, Wald's $\chi^2(1) = 4.26$, $p < .05$, OR = 4.02, all other ps $> .23$. Rules were more likely to be accurate in the extreme stereotypicality conditions compared with the conditions with less stereotype-relevant information (see Table 2). Confidence in the accuracy of the rule reflected a similar pattern, $F(4, 77) = 2.71$, $p < .05$, with participants in the extremely stereotype-consistent and -inconsistent conditions more confident compared with those in the moderate and neutral conditions (although post hoc tests showed significant differences only between the moderately inconsistent and the two extreme conditions).

Experiment 3 provided additional evidence that stereotypicality affects social category accessibility in a curvilinear fashion. Rule learning was superior in the extreme stereotypical and extreme counterstereotypical conditions, consistent with the stereotype-relevance hypothesis. Moreover, a graded relation between these variables was observed. Category learning was slowest in the neutral condition, faster in the moderate conditions, and fastest in the extreme conditions. In contrast with the novelty hypothesis and a marginal finding from Experiment 2, Experiment 3 indicated that category learning was equivalent in the extreme inconsistent and the extreme consistent conditions.

General Discussion

Our results show that category accessibility is moderated by the stereotypicality of targets' behavior. In two experiments, learning a category-based rule was easiest and fastest when participants were shown stereotype-consistent and stereotype-inconsistent behaviors compared to stereotype-irrelevant behavior. These results show that categories are likely to become accessible when individuals encounter stereotype-relevant behavior information.

This work represents an advance in both methods and theory. In terms of methods, this research introduces the ICAT as a tool for exploring spontaneous categorization processes. It allows the use of complex materials showing social targets performing realistic behavior in interactions with others. In contrast to existing methods, it permits the estimation of the relative speed of category activation. However, it is important to note that the ICAT is a measure of category accessibility in the context of an explicit instruction to search for features that differentiate the categories of stimuli. It cannot be assumed that the processes underlying the ICAT would occur in the absence of such prompting. Nonetheless, the ICAT does appear to be useful for assessing variables that affect categorization accessibility (e.g., the stereotypicality of behavior) within such a context.

Additional research is needed to better understand the exact processes underlying performance on the task. We assume that perceivers are actively attempting to "solve" the puzzle presented in the ICAT by testing various solutions, and stereotype-relevant behavior presumably facilitates generating correct hypotheses. We also suspect that people for whom gender is chronically accessible might generate gender-based rules more readily, and those rules might be easier or more difficult to test given the availability of stereotype-confirming versus-disconfirming information. However, we do not yet have direct evidence of different hypotheses that individuals generate when completing the ICAT, although we are currently conducting studies that do so.

Also, we have used the ICAT to study gender, a dimension that is readily noted and used in social information processing. The ICAT can be used to assess the accessibility of other important categories such as race, attractiveness, and weight simply by modifying the pictorial stimuli used. We can think of no a priori reason why this method would not be helpful for investigating a broad range of categories important in social judgment and behavior.

Implications for Stereotype Maintenance and Change

The current findings also speak to theories regarding stereotype change in response to stereotype-inconsistent information. Based on evidence demonstrating a lack of generalization from atypical exemplars to their social group, it has been argued that categories might not become activated on exposure to stereotype-inconsistent exemplars (e.g., Rothbart & John, 1985; Rothbart & Lewis, 1988). Our findings, however, show that the failure to generalize does not necessarily implicate the absence of category activation. In our studies, stereotype-inconsistent exemplars were just as likely to activate gender categories as stereotype-consistent exemplars. Why might exemplar-to-group generalization not occur for stereotype-consistent exemplars? We suspect that categorization of individuals who defy stereotypes initiates other processes such as biased attributions, creations of subtypes, and controlled activation of alternate categories. When perceivers encounter a person whose behavior contradicts stereotypical expectations, they notice the discrepancy between the category and the expected behavior. However, motivations to preserve existing beliefs will instigate processes that serve to maintain and protect existing category associations.

Will encountering stereotype-inconsistent behavior always prompt controlled processes to account for the unexpected information? We suspect that there are conditions where stereotypical behavior might play a different role, specifically in cases where category membership information is itself ambiguous or vague. In our studies, the stimuli were selected so that targets' sex was clear and unambiguous. What if, instead, the group membership of the target was vague? Under such conditions, behavioral information might be used to disambiguate categorization judgments, facilitating categorizations in a stereotype-consistent fashion. Hugenberg and Bodenhausen (2004), for example, showed that faces of ambiguous-race targets were more likely to be categorized as Black to the degree they exhibited hostility, particularly by individuals high in implicit anti-Black attitudes. Thus, stereotypicality of behavior can be used both to activate categories (when the discrepancy between behavior and categories is clear) and to infer categories (when category membership is vague or uncertain).

Notes

1. We also coded 8 of 10 and 9 of 10 sequentially correct responses as indicative of learning. This looser coding increased the amount and ease of learning in all studies, but the pattern of differences obtained across conditions was unaffected.
2. Because distribution-free procedures for factorial designs are not well developed, we followed the recommendation of Conover and Iman (1981) to convert the learning trials to ranks to allow use of parametric analyses.
3. Because of the small number of male participants, participant sex was not included as a variable in the analyses.

Acknowledgments

We express our thanks to Sara Menlow Wasserman, Carolyn Greene, Athena Masci, Rachel Shiovitz, Lauren Alfano, Jill Caruso, Beth Kachur, and Carrie Logerfo for their assistance in developing stimuli and data collection and to Lauren Aguilar for offering comments on an earlier draft of the manuscript. We are also indebted to Peter Balsam for his numerous helpful suggestions on this work.

Declaration of Conflict of Interest

The authors declared that they had no conflicts of interests with respect to their authorship or the publication of this article.

Financial Disclosure/Funding

This research was supported by a Third Millennium Foundation grant to the first author.

References

Allen, T. J., Sherman, J. W., Conrey, F. R., & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology*, *45*, 1081-1087.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437-449.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*, 763-778.

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, *29*, 43-74.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*, 123-152.

Cantor, N., & Mischel, W. (1979). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality*, *13*, 187-205.

Conover, W. J., & Inman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124-134.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. Reading, MA: Addison-Wesley.

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.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509-517.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*, 289-302.

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

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Quach, R. (2002). Mechanisms of source confusion and discounting in short-term priming 2: Effects of prime similarity and target duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1120-1136.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342-345.

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin & Review*, *1*, 56-72.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136-153.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*, 1155-1178.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*, 522-544.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275-287.

Locke, V., Macrae, C. N., & Eaton, J. L. (2005). Is person categorization modulated by exemplar typicality? *Social Cognition*, *23*, 417-428.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93-120.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*, 397-407.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology*, *33*, 471-489.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Reviews in Psychology*, *35*, 113-138.

Monaci, G., Menegaz, G., Susstrunk, S., & Knoblauch, K. (2004). Chromatic contrast detection in spatial chromatic noise. *Visual Neuroscience*, *21*, 291-294.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353-363.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues*, *41*, 81-104.

Rothbart, M., & Lewis, S. (1988). Inferring category attributes from exemplar attributes: Geometric shapes and social categories. *Journal of Personality and Social Psychology*, *55*, 861-872.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology*, *32*, 309-336.

Schneider, D. J. (2004). *The psychology of stereotyping*. New York, NY: Guilford.

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

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

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*, 3-21.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, *96*, 58-83.

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition*, *14*, 247-276.

Taylor, S. E., Fiske, S. T., Etcoff, N. J., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778-793.

Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R., & Houle, S. (1994). Novelty encoding networks in the human brain: Positron emission tomography data. *Neuroreport*, *5*, 2525-2528.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, *8*, 161-185.

Bios

Steven J. Stroessner is a Professor of Psychology at Barnard College, Columbia University.

Elizabeth L. Haines is an Associate Professor of Psychology at William Paterson University.

Jeffrey W. Sherman is a Professor of Psychology at the University of California, Davis.

Cara J. Kantrowitz is a graduate of Barnard College and is currently a Research Analyst at Insurance Archaeology Group, Ltd.