Title

Judging Compound Social Categories: Compound Familiarity and Compatibility as Determinants of Processing Mode

Journal

Social Cognition, 23(4)

ISSN

0278-016X

Authors

Groom, Carla J
Sherman, Jeffrey W
Lu, Lisa
et al.

Publication Date

2005-08-01

DOI

10.1521/soco.2005.23.4.291

Peer reviewed

Social Cognition, Vol. 23, No. 4, 2005, pp. 291-323

JUDGING COMPOUND SOCIAL CATEGORIES: COMPOUND FAMILIARITY AND COMPATIBILITY AS DETERMINANTS OF PROCESSING MODE

Carla J. Groom
KRC Research, London, UK

Jeffrey W. Sherman
University of California, Davis

Lisa Lu
Northwestern University

Frederica R. Conrey
Indiana University

Boukje S. C. Keijzer
University of Amsterdam

Three experiments tested the hypothesis that judgments about a group formed by two paired categories would rely on stored instances of individual category members (i.e., exemplars) in some cases, but not in others. Specifically, judgments of a relatively unfamiliar compound category (e.g., male elementary schoolteachers) were expected to rely on exemplars, whereas alternative sources of information, particularly abstract stereotypes, would be available for making judgments of a more familiar category (e.g., female elementary schoolteachers). Experiments 1 and 2 demonstrated support for these hypotheses. Experiment 3 ruled out the possibility that the differences in judgment strategy between the familiar and unfamiliar compound categories arose from the relative incompatibil-

This research was supported by NIMH grants R03 MH55037 and R01 MH59774. Correspondence should be addressed to Jeffrey W. Sherman, Department of Psychology, University of California, One Shields Ave., Davis, CA, 95616; E–mail: jsherman@ucdavis.edu.

ity of the two constituent categories (e.g., males and elementary schoolteachers), rather than familiarity. Implications for stereotype development and change are discussed.

If you were asked to explain what Mexican women were like, what information would you use to construct an answer? Would you be more likely to extrapolate from the memory of a particular Mexican woman that you know, or to draw on some kind of abstract stereotype of Mexicans, women, or Mexican women? Much social psychological attention has been devoted to whether people use exemplars or abstractions to make judgments about simple social categories, such as women or engineers (Hamilton & Sherman, 1994; Sherman, 1996; Sherman, Klein, Laskey & Wyer, 1998; Smith, 1990; Smith & Zarate, 1992). Yet, in the real world, people belong to multiple social categories, from gender to occupation to religion to age. Accordingly, perceivers must have a mechanism for inferring attributes of compound categories—that is, categories defined by the intersection of two or more constituent categories.

Little social psychological work has focused explicitly on how exemplar and abstraction strategies apply to judgments of compound categories. However, there has been closely related work seeking to understand how subcategories (e.g., grandmotherly types) are represented in relation to superordinate categories (e.g., elderly people). Two major models of social subcategory representation exist. To the extent that compound categories can be construed as subcategories, these models make relevant predictions about how judgments of compound categories will be formed.[1]

Brewer and her colleagues have suggested that subcategories are judged using information contained in abstract prototypes.

1. Technically, of course, compound categories can be construed as a type of subcategory in the sense that compounds are a logical subset of each of the constituent categories. However, it is important to bear in mind that logical and psychological category organization need not be the same (see Hampton, 1988). Thus, the compound "Harvard-educated carpenter," though logically a subcategory of both Harvard graduates and carpenters, may be perceived as quite distinct and dissimilar to both constituent categories, and would not be considered a subtype of either.

Brewer (1988) proposed a hierarchical model of superordinate–subordinate category relations. This model asserted that people partition the world along a few basic dimensions (e.g., age, sex, race), and psychologically nested within these are prototypes corresponding to more specific categories. These more specific categories need not be compound categories. For example, in an early study (Brewer, Dull, & Lui, 1981), Brewer explored the mental representation of the elderly, and presented data consistent with the notion that people possess three prototypic representations corresponding to three subcategories of the elderly: the senior citizen, the grandmotherly type, and the elder statesman. Nevertheless, some of Brewer's original examples did draw on classic compound categories, where she argued that the attributes of such categories would be tied to the abstract representation of one or another constituent category. For example, Brewer suggests that *woman doctor* is probably a subset of *woman*, whereas *man doctor* ("or more simply, doctor," p. 12) is probably a subset of *doctor*. She argues that because *man doctor* and *woman doctor* are subsets of different superordinate categories, different attributes will be ascribed to each subcategory. Thus, Brewer's model would predict that Mexican women would be judged via reference to a Mexican woman prototype, nested within either the superordinate category of women or that of Mexicans, though not all the features of the subcategories would necessarily be derived from the superordinates.

In contrast, Smith (1990) suggested that abstract information is not normally acquired for categories narrower than race, sex, etc. He argued that, as categories become more specific (e.g., sprinter vs. athlete), they are more likely to be understood in reference to past encounters with individual subcategory members. Thus, Smith's model predicts that exemplars are moderately likely to be used in judgments of Mexican women, very likely to be used to judge middle–aged Mexican women, almost certain to be used to judge middle–aged, lesbian Mexican women, and so on.

To our knowledge, no research has directly tested the extent to which knowledge of subcategories or compound categories is based on specific exemplars versus abstract knowledge.

COMPOUND FAMILIARITY AS A MODERATOR

Smith's exemplar model and Brewer's abstraction model could both be right, but under different circumstances. Specifically, exemplars may be used to judge relatively unfamiliar compound categories, whereas abstractions may be used more often in judgments of familiar compounds.

There is a good deal of evidence that social judgments are exemplar–based when the target is unfamiliar, but that exemplar use diminishes as familiarity increases. This developmental pattern applies to judgments of both self (in novel vs. familiar contexts) and other individuals (e.g., Klein, Loftus, Trafton, & Fuhrman, 1992; Klein & Loftus, 1993; Sherman & Klein, 1994). Most relevant is Sherman's work on the development of group stereotypes (1996; see also Sherman et al., 1998). He reported that exemplars were used only to judge novel groups comprising unfamiliar individuals. In one experiment, participants read a series of behaviors performed by different members of a novel university club. When asked to make trait judgments about this group, participants spontaneously retrieved these exemplar representations. However, with sufficient exposure to information about this particular group, evidence of exemplar–based inference disappeared. Sherman attributed this shift away from exemplars to the creation of a supplementary abstraction that was applied in preference to exemplar retrieval. Furthermore, if participants were told at the outset that the group members were all engineers or all priests, thereby supplying a pre–existing abstract stereotype, then stereotype–relevant group judgments (e.g., of intelligence for engineers) did not involve the recruitment of exemplars even when exposure to the group was minimal.

The suggestion of the present paper is that compound categories develop in much the same way, regardless of the familiarity of the constituent categories. For example, *Mexicans* and *women* are both familiar categories for which most people are likely to possess abstract stereotypes. However, the compound category *Mexican woman* may be sufficiently unfamiliar to many perceivers that their judgments will draw on experiences with individual Mexican women, if such experiences are available in memory. As experience with Mexican women grows, however, an abstract

stereotype of the group should develop, diminishing the role of exemplars in group judgments.

FAMILIARITY AND COMPOUND CATEGORIES: PAST RESEARCH

Though there is a considerable body of research examining the nature of judgments about compound categories, very little of it has examined the role of familiarity (for reviews, see Medin & Rips, 2005; Rips, 1995). The reason is because most of this research has been concerned exclusively with the nature of compound judgments at zero familiarity, when participants have never encountered a single compound exemplar (the same applies to work on "ad hoc" categories; Barsalou, 1983). At zero familiarity, the key questions surround the means by which people combine knowledge from two concepts that have never been combined to produce new knowledge. In this case, knowledge of the compound must be derived "intensionally," from the constituent categories alone. The presence of prior knowledge of compound exemplars has been seen primarily as an impediment to gaining a clear understanding of these conceptual combination processes (e.g., Medin & Rips, 2005; Rips, 1995).

Nevertheless, the handful of existing studies addressing judgments of compound social categories provides general support for the notion that familiarity influences the choice of judgment strategies. In a study focusing on the strategy of causal reasoning, Kunda, Miller, and Claire (1990) asked participants to describe features that would be expected for members of various compounds. Compounds that were rated as more surprising were associated with more frequent mentions of causal antecedents. For example, blind marathon runners, a highly surprising compound, were ascribed "much courage and ability to overcome their handicap" (p. 559). Participants seemed to generate narratives that made the compound more coherent and plausible. Surprising compounds were also more frequently associated with emergent attributes (i.e., properties mentioned for the compound but not for its constituent categories). Less surprising compounds such as Harvard–educated lawyers elicited fewer indications of

causal reasoning. Because subjective feelings of surprise are likely to be highly negatively correlated with familiarity, this study provides evidence that familiarity influences the strategy used to predict the attributes of social compounds. In particular, it suggests that constituent superordinate categories have less influence on judgments of unfamiliar than familiar compounds.

A study by Hastie, Schroeder, and Weber (1990) provided additional evidence that familiarity influences the outcome and process of compound category judgment. Attribute–listings and trait–ratings were obtained for category compounds and their respective constituent categories. Compared to familiar compounds such as *male mechanic*, unusual compounds such as *Republican social worker* yielded more emergent attributes, and also more outside ratings (i.e., ratings falling outside the range defined by the constituent category ratings on the same trait), suggesting less reliance on constituent categories for these judgments. Hastie et al. obtained preliminary evidence of judgment processes by asking their participants to introspect about the strategies that they had used to generate trait ratings. Although in the case of both familiar and unfamiliar compounds the dominant observation was that the ratings had been derived from stereotypes of the constituent categories, there were more reports of "complex" processes when judging unfamiliar compounds. Complex processes included, for example, causal arguments (e.g., a woman holding down a "man's" job must be a certain kind of woman). A very small proportion of responses referred to generalization from exemplars of either the compound or an analogical category (e.g., a female construction worker in lieu of a female mechanic).

Although exemplar–based strategies attracted little attention in these studies, this is most likely because the unfamiliar compounds were extremely unfamiliar. If the undergraduate participants possessed no stored exemplars for a compound, such as Harvard–educated carpenters or Republican social workers, it obviously would be impossible for them to have used an exemplar–based judgment strategy.

The literature on non–social concept combination has examined the role of exemplars more directly. Following the frequently unsuccessful attempts to find a way to predict the properties of a compound from the properties of constituent categories (e.g., Co-

hen & Murphy, 1984; Smith & Osherson, 1984), several researchers have pointed out that people's knowledge of compound members is one of the most likely explanations for this "non–compositionality" of compound categories. Hampton (1988) ascribes a good deal of non–compositionality to knowledge formed "extensionally" from compound members, rather than intensionally from constituent prototypes. Along these lines, Gray and Smith (1995) explored the phenomenon of "typicality reversals," in which a property is more salient for a compound than for constituent categories (e.g., *metal* is rated as more typical than *wooden* as a property of *spoon*, but the reverse is true when rating *large spoon*). They found support for the hypothesis that typicality reversals tended to occur when the reversed property of the compound (e.g., *wooden* in the case of *spoon*) was also salient to participants who were simply asked to retrieve exemplars of the compound, thereby implicating the role of exemplars in such reversals. Gray and Smith also manipulated the familiarity of their target compounds and found that unfamiliar compounds were less likely than more familiar compounds to be related to instance properties. However, as in the case of the social compound studies, the unfamiliar compounds were extremely unfamiliar (e.g., square ball, triangular book), and do not address the early stages of category development when the first contact with compound members is being made.

In summary, there is already evidence that exemplars may be used when deduction is unavailable, and that familiarity affects the choice of strategy for understanding compounds. The present research aimed to bridge the gap by directly examining the role of exemplar processes in judgments of compound social categories.

HYPOTHESIS

We hypothesized that people would make exemplar–based trait judgments of social compound categories, consistent with Smith's model, when those compounds were relatively unfamiliar. Based on the work of Sherman and Klein, a developmental shift in judgment strategy was predicted as familiarity with the compound increased. For relatively familiar compounds, we ex-

pected that perceivers would possess an abstraction specific to that compound, and prefer to deduce traits from this abstraction despite the availability of exemplars in memory. Judgments of familiar compound categories would therefore be more in line with Brewer's model.

We chose the compound categories *male elementary schoolteacher* and *female elementary schoolteacher* to represent the unfamiliar and familiar compounds, respectively. These compounds were specially selected because they fit several criteria necessary to provide a pure test of the hypothesis. First, these combinations do not differ in the familiarity of the constituent categories (i.e., males and females are equally familiar, and elementary schoolteacher is common to both compounds), which ruled out one important potential confound. Our developmental hypothesis assumed that familiarity with members of the compound itself would be critical for determining judgment strategy, regardless of the familiarity of the constituents. Second, it was important to control for the specificity of the compounds because one claim of Smith's model was that exemplar usage would be strongest for more specific compounds. These compounds were equally specific (i.e., both comprise gender + elementary schoolteacher), ruling out the possibility that specificity could explain any difference in judgment strategies that we observed. Third, we required that the unfamiliar compound should be one for which the vast majority of our participants would have a few exemplars, but only a few, available in memory. Male elementary schoolteachers fit the bill. Our participants could reasonably be assumed to have had considerable experience with elementary schoolteachers, comprising of mostly females but a few males. Thus, participants should be able to appeal to a female elementary schoolteacher abstraction in order to determine the attributes of female elementary schoolteachers, but would need to draw on exemplars to judge male elementary schoolteachers.

EXPERIMENT 1

OVERVIEW AND PREDICTIONS

The first study sought to provide an initial demonstration of the predicted asymmetry in strategies for judging compound catego-

ries using a methodology comparable to those used in previous studies of social compounds. Participants were assigned to judge either male or female elementary schoolteachers, and were asked to make two sets of trait judgments using identical scales. First, they were asked about the properties of the compound category as a whole (i.e., to rate male or female elementary schoolteachers in general). Participants then recalled a particular male or female elementary schoolteacher from their own experience, and made a second set of judgments on the same trait dimensions, but focusing on this particular person. Judgments about the individual group member should resemble ratings of the entire group to the extent that perceivers rely on memories of particular individuals to construct inferences about the compound. Correspondence between the two sets of ratings should diminish to the extent that perceivers draw on alternative sources of information about the group, particularly abstract representations. The extent of correspondence was measured by calculating the difference between group and exemplar ratings on each trait for each participant. We inferred that the smaller the absolute group–exemplar differences, the more likely it was that the exemplar had been used to produce the group judgments. Thus, we expected that the differences would be smaller for judgments of male targets than for judgments of female targets.

We also were interested in whether participants in the male target condition would be more likely to be consciously aware of using exemplars when making group judgments, relative to participants in the female target condition. Although exemplar use could occur implicitly, positive evidence of such a phenomenological difference would provide supplemental support for our hypothesis, subject to the acknowledged limitations on people's ability to report the bases for their judgments (e.g., Nisbett & Wilson, 1977).

METHOD

Participants. Fifty–four undergraduate volunteers were recruited by approaching students at random inside the Northwestern University student center.

Materials and Procedure. Half of the participants were asked to rate female elementary schoolteachers on five traits (energetic,

wordy, extraverted, friendly, temperamental), then to repeat the task with reference to a specific female elementary schoolteacher that came to mind. The rating scales ranged from 1–10 (e.g., 1 = *very succinct*, 5 = *somewhat wordy*, 10 = *long winded*). The remaining participants were given the same rating scales to make judgments of male elementary schoolteachers in general, and then to rate a specific male elementary schoolteacher that they could recall. All participants were then asked whether they had been able to think of a particular person, and whether they had thought about that person when making the group judgments. Finally, participants were probed for suspicion and debriefed.

RESULTS

Only four participants failed to recall a specific teacher, and their data were excluded. For each remaining participant, a difference score on each of the five traits was created by subtracting the ratings of the specific individual from ratings of the group. The absolute values of these difference scores were submitted to a 2 (target: group vs. individual) × 5 (trait: energetic vs. wordy vs. extraverted vs. friendly vs. temperamental) ANOVA with repeated measures on the second factor.[2] This analysis produced only the predicted main effect of target. The mean absolute difference score when female elementary schoolteachers were the target (M = 1.82) was found to be greater than the mean absolute difference score when the targets were male (M = 1.11), F (1, 48) = 5.94, $p <$ 0.05. These results demonstrate that exemplar ratings correspond more closely to general group ratings in the case of male as compared to female elementary schoolteachers.

The proportion of participants reporting that they had thought of the same person during both judgments was also much greater in the male target condition (80%) than the female target condition (16%); F (1, 48) = 33.39, $p < .01$. This reinforces the claim

2. Note that we make no predictions about the directions of the differences in ratings of groups and individuals (i.e., individuals may be judged more or less extremely than the groups). We predict only that the differences (regardless of direction) will be larger in the case of male than female targets. Thus, it is appropriate to analyze the absolute values of the differences in ratings.

that there is an asymmetry in judgment processes for male and female elementary schoolteachers, and also indicates that perceivers have some conscious access to the type of inference strategy that they are using. Furthermore, participants' beliefs corresponded to actual exemplar usage; the average difference between specific and general ratings was indeed smaller for participants who believed they were relying on the same exemplar ($M = 0.98$) compared to those who did not ($M = 1.92$), $F\ (1, 48) = 11.71$, $p < 0.01$.

DISCUSSION

As predicted, judgments of the group and judgments of a group member are based on representations that tend to overlap more in the case of male elementary schoolteachers than in the case of female elementary schoolteachers. This is consistent with the hypothesis that perceivers are more likely in the case of male rather than female elementary schoolteachers to use exemplars to infer features of the compound category as a whole. According to our hypothesis, the pairing of male and elementary schoolteacher is sufficiently unfamiliar to preclude the availability of a specialized abstraction for the compound. The familiar pairing of female with elementary schoolteacher does not pose this problem, and it appears likely that participants are able to rely on abstract knowledge of the compound to form judgments.

Phenomenological evidence supports our suggestion that exemplars were being used more often during judgments of male than female group targets. Participants in the male target condition were more likely to report having had the same individual in mind for both the individual and group judgments. Furthermore, those participants reporting this experience did tend to make more similar ratings of the individual and group targets. It seems that when group judgments are exemplar–based, perceivers may have explicit access to that process.

In addition, Experiment 1 confirmed our assumption that the vast majority of our participant population has exemplars of male elementary schoolteachers, as well as female elementary school-

teachers, available in memory. Two out of twenty–seven (7%) participants failed to recall a specific male teacher, and two out of twenty–seven failed to recall a specific female teacher.

EXPERIMENT 2

The conclusions we can draw from Experiment 1 are limited by various aspects of the methodology. First, the comparison of group and individual trait ratings is not an ideal index of exemplar usage. Despite subjective reports to the contrary, we cannot rule out the possibility that close correspondence between the two ratings arose not from the use of exemplars to make group judgments, but rather from the opposite process (i.e., the use of group abstractions to infer the features of a particular individual). For example, when a perceiver is asked to think of a familiar male elementary schoolteacher and describe his traits, he/she may use a male elementary schoolteacher abstraction to fill in details about that person that have been lost, never observed or stored, or simply require effort to retrieve. This process seems particularly likely using a closed response format like ours, because participants were expected to provide ratings for every trait, regardless of how confident they were about their information. Furthermore, this process also seems particularly likely in the case of male targets in Experiment 1. People are likely to have far fewer male elementary schoolteachers than female elementary schoolteachers stored in memory, and therefore fewer candidates available in the male target condition. Assuming that participants generally retrieve the most personally salient and well–remembered exemplar from the available candidates, then the average exemplar retrieved in the female target condition will be better remembered than the average exemplar retrieved in the male target condition.

A second and related problem is that the overlap between the listed exemplar and the group rating is likely to be smaller to the extent that many different exemplars are retrieved for the group judgment. Thus, if more female than male teachers were retrieved for the group judgments, then the influence of any single exemplar would have been smaller in the case of the female teachers,

resulting in a greater difference between the group and individual judgments for female than for male targets.

Another difficulty is that the measure is, at best, relative, allowing us to rank target groups in terms of their reliance on exemplars for judgment, but not to determine whether judgments of lower–ranked groups recruit exemplars at all. Thus, the results of Experiment 1 could not tell us whether exemplars are never used during group judgments of female elementary schoolteachers or whether they are simply used less than when judging male elementary schoolteachers. Experiment 2 resolved these problems by using the same target groups but a different methodology that provided a more absolute measure of exemplar retrieval.

MEASURING EXEMPLAR ACTIVATION: A PRIMING PROCEDURE

To measure exemplar activation in Experiment 2, we used a priming measure that has been validated in several studies on self–, other–, and group–judgments (Klein, Babey, & Sherman, 1997; Klein & Loftus, 1993; Sherman, 1996; Sherman & Klein, 1994). In the version of the procedure used in this experiment, participants first read some behaviors that were supposedly reported by a particular group. Participants then performed a critical two–task sequence. The second task was always a recall task, which required participants to recall a specific incident in which a member of the target group behaved in accordance with a particular trait (e.g., "Remember a specific incident in which a member of the group behaved in a kind manner"). The initial priming task was either a control task or a group judgment task. The group judgment task asked participants to decide whether the target trait was generally applicable to the group that the participants had read about (e.g., "Does the word *kind* describe the group?"). The control task asked participants to define the target trait (e.g., "Think of the meaning of the word *kind*"), and served as a baseline. The type of priming task was the key independent variable. Time taken to perform the recall task served as the main dependent measure. The recall latency was measured by asking participants to press a

space bar when they had successfully retrieved a specific behavioral incident.[3]

If judgments are based on information about group exemplars, then it should take less time to recall a group behavior following a group judgment task than following a control task. This is because group behaviors will already have been activated during performance of the group judgment task but not of the control task (i.e., the behavior will have been primed). Past research has supported the assumption that the trait definition control task does not activate exemplars (Klein et al., 1997). If judgments are based on abstract information about the group or on other non–exemplar inference strategies, then the judgment task should not facilitate recall compared to the control task. An advantage of this task is that it is sensitive to both explicit and implicit exemplar activation. Even if exemplars have not been explicitly retrieved, but have been activated implicitly, they should subsequently be retrieved more quickly. A related advantage of this measure is that it does not depend on participants having conscious access to their judgment strategies.

OVERVIEW AND PREDICTIONS

Experiment 2 assessed the extent to which judgments about groups of male and female elementary schoolteachers were reliant on exemplars. After reading about a fictitious group of male or female elementary schoolteachers, participants were trained to perform the priming task described above. The experiment was a 2 (initial task: group judgment vs. control) × 2 (target sex: male vs. female teachers) between–subjects design. We predicted that exemplar usage would be found in the male target condition but not the female target condition. As such, performing the judgment task would lead to faster recall compared to the control condition in the male but not the female target group.

3. We did not request that participants report their responses during the experimental trials; rather, we instructed them to generate responses to the task questions in their heads. See Klein and Loftus (1993) for a detailed discussion of the rationale for this procedure, and research demonstrating the efficacy of the technique.

METHOD

Participants. Seventy Northwestern undergraduates participated for course credit.

Materials and Procedure. Participants read ten statements purportedly collected at a conference of either male or female elementary schoolteachers, and were asked to form an impression of the group of teachers about whom they read. Participants were told that a different teacher had reported each behavior, but that in the interests of anonymity no names were presented with the statements. Of the ten statements, two statements were behaviors that were pilot tested to be moderately indicative of kindness (e.g., "he/she let another car into the line of traffic," "went to the grocery store with his/her nephew"), four were behaviors irrelevant to kindness (e.g., "had orange juice for breakfast") and four were demographic facts (e.g., "was born in Minnesota"). Previous work (Sherman, 1996) had found that this number of behavior statements was insufficient for participants to develop an abstract representation of a novel group, and that the group would be judged with reference to those exemplars unless participants could apply an existing relevant stereotype.

After reading the stimuli, participants were trained to perform the judgment, control and recall tasks. In order to provide a context for training, participants were asked to focus on a group of their friends. All three tasks were explained, and practice was given using a series of two–task sequences. At the start of each sequence, a computer displayed the task instruction ("DEFINE," "DESCRIBE" or "RECALL"), accompanied underneath by a target trait (e.g., "Musical"). Participants were asked to perform the task mentally, and to indicate completion by pressing the space bar. For example, if they saw the instruction "DESCRIBE" above the trait "Thrifty," they were required to decide whether their group of friends could be described as thrifty, and press the space bar when they had decided. A second task type followed, accompanied by the same trait, and again a bar–press reflected task completion time. A fresh sequence then began, using a different trait and any two of the three tasks. None of the traits used during this training phase were relevant to the key trait, kindness, that was used in the critical two–task sequence that constituted the primary dependent measure. The importance of actually com-

pleting the task in their heads prior to pressing the space–bar was impressed upon the participants.

When training was complete, participants were asked to perform the task with reference to the particular group of teachers that they had read about earlier in the experiment. They were randomly assigned to receive either a control–recall or judgment–recall sequence. The trial began with one of two cues for the priming task appearing on the screen: "DESCRIBES MALE [FEMALE] TEACHER GROUP" (judgment task) or "DEFINE" (control task). Two seconds later, the trait *kind* appeared below the task cue. The cue and the stimulus trait remained on the screen until the participant pressed the space bar to indicate that she/he had completed the judgment or define task. After a two–second pause during which the screen was blank, the cue for the recall target task ("RECALL MALE [FEMALE] TEACHER GROUP") appeared on the screen above the same trait, *kind*. Once again, the screen did not change until participants pressed the space bar to show that they had completed the recall task. The computer recorded the time required to perform the recall task, which constituted the dependent measure. After this single two–task sequence was completed, the trial was over. Before debriefing and dismissal, participants wrote down the particular behavior that they had recalled for the recall task, as a compliance check.

RESULTS

Five participants were excluded from the analyses for failing to recall a kindness–related behavior, according to the compliance check. Four further participants were excluded as outliers based on their recall latencies. Outliers were defined as responses 2.5 standard deviations above the mean. Figure 1 shows the recall task latencies broken down by priming task type (judgment vs. define) and target group (male vs. female). The predicted interaction was found between target type and task type $F(1, 58) = 4.34$, $p < .05$. In the case of male targets, recall of a trait–related behavior was faster after judging whether the trait described the group ($M = 4579$ ms) than after defining the trait ($M = 6247$ ms), $F(1, 30) = 4.93$, $p < .05$. However for female targets, there was no significant

difference in recall time between the judgment condition (M = 5549 ms) and control condition (M = 5011 ms), $F(1, 28) = 0.52$, ns.[4]

DISCUSSION

Consonant with our predictions, the task of assessing the kindness of a group of male elementary schoolteachers involved the activation of specific exemplar representations, thereby facilitating subsequent recall of exemplars relative to a baseline task. In contrast, thinking about the kindness of a group of female elementary schoolteachers did not facilitate recall, and therefore appears not to have evoked exemplars. These data are consistent with the idea that the features of compounds can be judged solely on the basis of the abstract knowledge specific to the compound, provided that those compound categories are relatively familiar. Where compound categories are relatively unfamiliar, an appeal to exemplars may be made.

EXPERIMENT 3: FAMILIARITY VERSUS COMPATIBILITY

Experiment 2 provided strong evidence that people draw on exemplars to make judgments about male but not female elementary schoolteachers. This suggests that the familiarity of the compound affects the likelihood that exemplars will be used during judgment. However, there remained a possible alternative ex-

4. It is worth noting that the crucial comparisons are those within a particular target sex condition. In the male target condition, it was predicted that trait judgments of the group would be based on activated behaviors. Therefore, recall should be faster following a judgment task than following an initial control task. In contrast, in the female target condition, it was predicted that group judgments would not be based on exemplars. As a result, it should take equally long to recall a behavior following either of the two initial tasks. However, we do not necessarily expect that a male behavior will be retrieved more quickly following a judgment task than will a female behavior. This is because at baseline (i.e., following the define task), female behaviors were retrieved more quickly than male behaviors. This replicates the common finding that retrieving information about familiar and abstract targets is faster and easier than is retrieving information about unfamiliar targets, both at the individual (e.g., Sherman & Klein, 1994) and group level (e.g., Sherman, 1996). Thus, the key prediction in the current research concerns the effect that a prior judgment task has on the subsequent retrieval of a group behavior. A prior judgment should facilitate, below baseline, retrieval of male but not female behaviors.

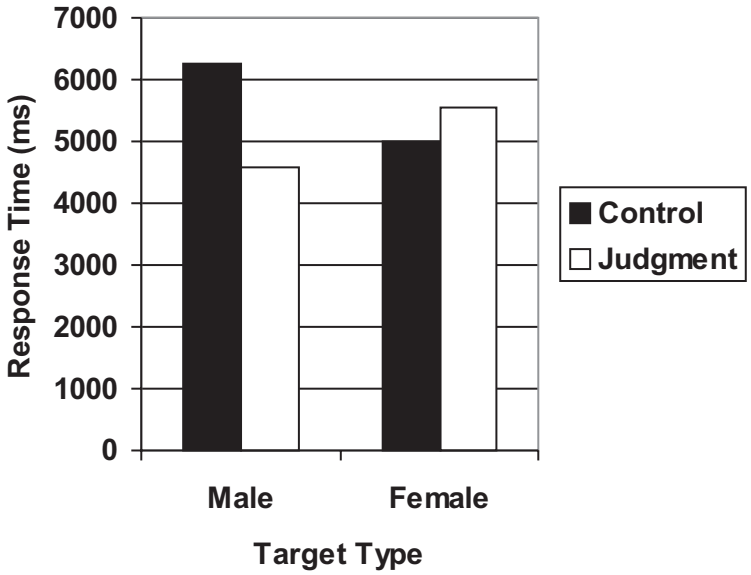

FIGURE 1. Recall latencies as a function of target type and initial task, Experiment 2.

planation for the effect: the choice between exemplar and abstraction strategies could depend on the compatibility of the constituent categories in the compound, rather than their familiarity, per se. The stereotype of elementary schoolteachers seems certain to overlap more with the stereotype of females than with the stereotype of males. It may be the case that when judgments of compound categories are required, the perceiver first defaults to the abstract stereotypes of the constituent categories, and only if these constituent stereotypes produce inconsistent inferences is it necessary to make a more effortful appeal to exemplars. This may explain why judgments of male but not female teachers are exemplar–based.

The same confound existed in the earlier study by Hastie et al. (1990). Compounds such as female mechanic are both incompatible and unfamiliar, whereas compounds such as male mechanic are both compatible and familiar. Indeed, Hastie et al. present their data alongside a model of social compound judgment that

portrays compatibility as responsible for the judgment strategy differences they observed. Many cognitive researchers share these authors' intuitions. A strikingly similar model has been offered by Thagard (1997) to explain conceptual combination in general, going beyond the social domain. Both Hastie et al. and Thagard proposed that perceivers' initial and preferred strategy would be for a straightforward reconciliation of the constituent category representations. This would probably occur through some process of averaging the values possessed by each constituent category on the attribute dimension to be judged (e.g., Anderson, 1981). A stable compromise representation would be the final product, from which deduction would be easy. In the case of female elementary schoolteachers or other relatively compatible compounds, a reasonable compromise should be possible because, on most dimensions, the stereotypes overlap. However, in the case of relatively incompatible constituents such as male elementary schoolteachers, the best compromise is frequently not good enough. For example, when asked to rate the kindness of male elementary schoolteachers, the predictions of each constituent category differ too strongly to make a compromise value satisfactory for either (elementary schoolteachers in general being stereotypically far kinder than males in general). Under such circumstances, the perceiver would switch to what Thagard termed "incoherence–driven processing." This would include exemplar–based induction as well as more complex, creative strategies such as causal reasoning and analogy.

The familiarity–compatibility distinction is not a trivial one. Although they should certainly be highly correlated in the real world, the two accounts represent quite different interpretations of what a compound category is, especially the extent to which compounds depend on their constituents and exemplars. Compound category formation from a compatibility perspective is essentially a combinatorial process, probably performed on an ad hoc basis, depending on the judgment required. This type of model was designed, primarily, to account for the creation of knowledge of novel compounds from known constituents at zero familiarity with the compound (e.g., Medin & Rips, 2005; Rips, 1995). This account portrays inter–category relationships as a fairly rigid hierarchical structure of superordinate and subordi-

nate categories working together to store knowledge efficiently. Only when combination is simply not possible, are attempts made to achieve reconciliation through reasoning, analogy, or exemplar–based induction. Of course, exemplar–based induction may occur only if compound exemplars are known. Thus, combinatiorial models are concerned primarily with intensional combination processes (i.e., how compounds are derived from general knowledge of their constituents), and specific exemplars belonging to the compound category play a rather minor role.

Conversely, compound category formation from a familiarity perspective is a developmental, continuous process that depends heavily on information gleaned from exemplars specific to the compound. Early in the developmental sequence, exemplars are directly retrieved for judgment purposes. Over time, the information extracted from these exemplars is used to develop an abstract representation specific to the compound category. Inter–category structure is much more fluid from this perspective, as superordinate constituents are not assumed to have any direct influence on the representation of compound categories (e.g., knowledge of male teachers may be based entirely on known male teachers, and not at all on prior concepts of males or teachers).

Fortunately, the two accounts do make different predictions in some cases, and it is possible to conduct an empirical test between them. Consider the case where perceivers are asked to rate male and female elementary schoolteachers on a dimension along which males, females and elementary schoolteachers are unlikely to differ, such as intelligence. A compatibility–based model would predict that there would be no need to appeal to exemplars for either compound, because the abstract stereotypes of male, female and elementary schoolteacher all concur. In contrast, a familiarity–based model would predict that male elementary schoolteachers still would be judged via exemplars because no specialized abstraction is yet available. On the other hand, perceivers are likely to have had sufficient opportunities to observe the intelligence of female elementary schoolteachers, and to have abstracted the central tendency.

A pilot test confirmed that, in ratings of intelligence but not kindness, the categories of *males* and *elementary schoolteachers*

were equally as compatible as were *females* and *elementary schoolteachers*. Thirty–six undergraduates from the participant pool at Northwestern University each made simple trait ratings (1 = *not stereotypical*, 10 = *very stereotypical*) of one of the following five groups: females, males, female elementary schoolteachers, male elementary schoolteachers, or elementary schoolteachers. On a minority of traits, including intelligence, the ratings for elementary schoolteachers were equally similar to those for males and females. The average absolute difference between males and elementary schoolteachers ($M = 1.05$) was very similar to that between females and elementary schoolteachers ($M = 1.22$), $F(1, 35) = 0.74$, ns. Kindness on the other hand did discriminate between the two compounds in the anticipated manner. The average absolute difference between males and teachers on ratings of kindness ($M = 2.36$) was much larger than that between females and teachers ($M = 0.86$), $F(1, 35) = 49.74$, $p < .01$.

Experiment 3 retained the same task and target compounds as Experiment 2, but substituted intelligence rather than kindness as the target dimension. This approach not only offered the possibility of ruling out compatibility as an explanation for our pattern of findings, but also extends the existing literature in an important methodological respect. Whereas past studies have systematically manipulated the composition of target compounds, less attention has been devoted to the choice of attributes to be judged, often relying on aggregates of trait ratings that mix dimensions on which constituent categories do and do not agree. For example, the 15 bipolar scales used by Hastie et al. (1990) were the same for all the compounds rated, and included such diverse traits as lower class, imaginative, ambitious, and warm. Indeed, Experiment 1 in the present paper selected traits with an eye to achieving and aggregating a diverse set of judgments, rather than their implications for constituent category overlap. Paradigms that ask participants to freely generate attribute lists cannot, by definition, control or manipulate the feature dimensions for which judgments are produced. To our knowledge, then, this is the first study of compound categories that highlights the attributes on which the compounds are judged, as well as the properties of the compounds themselves.

OVERVIEW AND PREDICTIONS

The primary purpose of Experiment 3 was to rule out compatibility as an alternative account of strategy choice during judgment of social compounds by replicating the pattern of results found in Experiment 2, using intelligence rather than kindness as the target trait. Thus, Experiment 3 used exactly the same priming procedure as Experiment 2, but replaced the kind behaviors and kindness–related tasks with intelligent behaviors and intelligence–related tasks. Once again, the procedure allowed us to detect whether exemplars were invoked when judging whether the target trait described the teacher group, and we predicted that performing the judgment task would lead to faster recall, compared to the control condition, in the male but not the female target group.

METHOD

Participants. Eighty–nine Northwestern undergraduates participated in the experiment for course credit.

Materials and Procedure. The procedure was identical to that used in Experiment 2, but the two key behaviors related to intelligence ("does the crossword puzzle in Sunday's paper," "read some novels over the summer") rather than kindness. The priming tasks involved either defining the trait *intelligent* or deciding whether *intelligent* described the teacher group that had been read about earlier. The dependent measure was the time taken to recall a behavior that a teacher had performed related to intelligence.

RESULTS

The data from four participants were excluded as outliers based on their recall latencies. As in Experiment 2, outliers were defined as responses over 2.5 standard deviations above the mean. Figure 2 shows the recall task latencies broken down by priming task type (judgment vs. define) and target group (male vs. female). The pattern found in Experiment 2 was replicated; recall facilitation following the judgment compared to the control task was

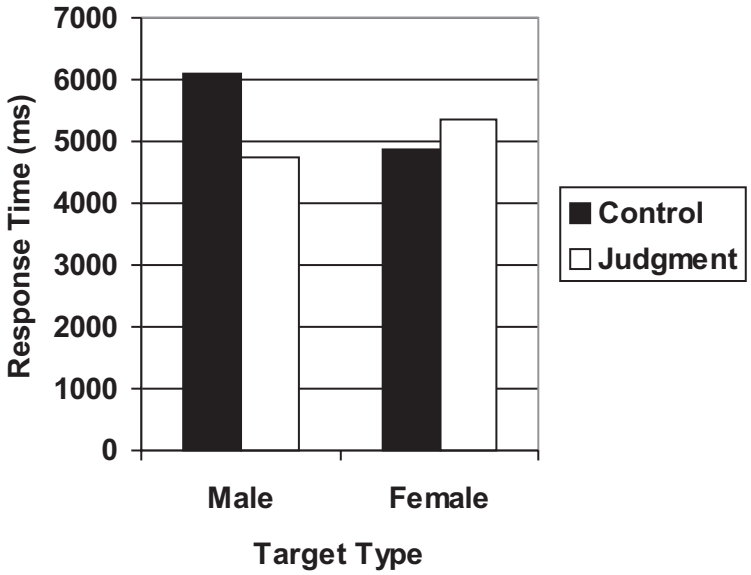

FIGURE 2. Recall latencies as a function of target type and initial task, Experiment 3.

found in the case of male targets but not female targets. A 2 (target type) × 2 (task type) ANOVA found the predicted interaction, F (1, 81) = 3.71, p = .058. In the case of male targets, recall of a trait–related behavior was marginally faster after judging whether the trait described the group (M = 4749 ms) than after defining the trait (M = 6098 ms), F (1, 30) = 4.93, p < .09. For female targets, latencies in the judgment condition (M = 5365 ms) were not faster than those in the control condition (M = 4864 ms), F = 0.74, ns.

DISCUSSION

Consistent with a familiarity rather than a compatibility–based account of compound category judgment, the results showed that judgments of a group of male elementary schoolteachers in-

volved recourse to exemplars, whereas judgments of a group of female elementary schoolteachers did not. Using judgments of intelligence, this study replicated the pattern predicted by our hypothesis and found using kindness judgments in Experiment 2. Because males and females are perceived to be equally similar in intelligence to elementary schoolteachers, this finding rules out the possibility that the asymmetry in exemplar–based inference is limited to traits on which males are less similar to elementary schoolteachers than are females to elementary schoolteachers. Exemplars were used in judgments of male elementary schoolteachers even though deductive inferences from the constituent category abstractions would have concurred in this case. Given that simple deduction was available, and yet participants still exerted the effort to retrieve exemplars, these data are difficult to reconcile with a view of compound category judgment as necessarily an intensional, combinatorial process. We propose that these results are best characterized in terms of a process that selects a strategy for inferring the features of a compound on the basis of the familiarity of the compound category in its own right.

BLOCKED Z–SCORE ANALYSIS

To attain a clearer picture of the overall strength of the influence of familiarity on exemplar use, we conducted a blocked meta–analysis in which the raw data from Experiments 2 and 3 were converted to z–scores within each experiment, and then combined into an overall ANOVA, with Experiment (2 vs. 3) as a between–subjects factor (Rosenthal, 1991). In this analysis, only the predicted two–way interaction between target type and task type was significant, F (1, 139) = 8.01, $p < .01$. For male targets, behaviors were recalled significantly more quickly following a judgment than a control task, F (1, 71) = 8.55, $p < .01$. For female targets, behaviors were recalled equally quickly following judgment and control tasks, F (1, 72) = 1.39, $p > .25$. These results provide strong support for the role of familiarity in determining the means by which judgments of compound social categories are achieved.

GENERAL DISCUSSION

Because people can be categorized simultaneously along a variety of dimensions, perceivers require a strategy for making inferences about members of compound social categories. One possible strategy is to make inductive judgments using exemplars, that is, memories of individual compound members. An alternative strategy might be to employ deduction based on an abstract stereotype of the compound. We used four existing models of social category judgment to derive predictions about how perceivers would judge compound categories. Though not directly concerned with compound category development, per se, Brewer's (1988) hierarchical model suggests that judgments would be based on abstract knowledge about the compound and of its constituent categories. Smith's (1990) exemplar model would predict that exemplars would be used to judge most compounds. A third class of models based on compound compatibility suggests that exemplars would be used only when the two constituent categories comprising the compound are incompatible with one another (e.g., Hastie et al., 1990; Thagard, 1997). Finally, based on the developmental mixed model (e.g., Klein & Loftus, 1993; Sherman, 1996), we hypothesized that judgments would be exemplar–based when the compound was relatively unfamiliar. Conversely, we expected that judgments of familiar compounds would not be based on exemplars because abstract knowledge should already be available in memory.

Three experiments tested these competing hypotheses by assessing perceivers' exemplar usage during judgments of two particular social compounds: male elementary schoolteachers and female elementary schoolteachers. These compounds differed in familiarity, but were similar in other key respects: most importantly, the specificity of the compound (gender + teacher) and the familiarity of the constituent categories (males and females are equally similar).

EVIDENCE OF FAMILIARITY–BASED DIFFERENCES IN EXEMPLAR USAGE

Experiment 1 showed that group judgments of male elementary schoolteachers more closely resembled judgments of an individ-

ual, accessible male teacher than did judgments of female elementary schoolteachers resemble judgments of individual female teachers. In addition, compared to female elementary schoolteachers, group judgments of male elementary schoolteachers were more likely to be accompanied by a subjective sense of retrieving a particular exemplar. These results suggest that judgments of male but not female teachers were made in reference to particular individual teachers.

In Experiments 2 and 3, participants learned about a limited number of male or female elementary schoolteachers. Experiment 2 demonstrated that judging the kindness of this group of male elementary schoolteachers did invoke exemplars of that group, whereas judging the kindness of the female group did not. Experiment 3 replicated the findings of Experiment 2 using intelligence judgments, a trait on which males and females are perceived to be equally similar to elementary schoolteachers. This replication demonstrates that the asymmetry between these compounds is not limited to those traits on which the stereotype of males is less compatible with the stereotype of elementary schoolteachers than is the stereotype of females. Such a finding is inconsistent with a judgment strategy that is determined solely by the compatibility of the constituent categories in the compound, and is in accordance with the hypothesis that familiarity with the compound itself is the key variable.

ABSTRACTION VS. EXEMPLAR MODELS: EVALUATION AND RECONCILIATION

A familiarity–based account of compound judgment qualifies the predictions that follow from both the abstract prototype theory of Brewer et al. (1981) and Smith's (1990) exemplar–based account. Our targets were equally specific, each containing identical occupational information and equivalent gender information. Whereas a prototype model would regard both compounds as subcategories, and postulate a similar, abstract representational basis for both categories, exemplars were consistently used during judgments of the male elementary schoolteacher compound, but not the female elementary schoolteacher compound. The data

also failed to support two claims of the exemplar model. First, in contrast to the prediction that exemplars are normally called upon for any categories narrower than race, sex, and so on, available exemplars were not called upon to judge female elementary schoolteachers. Moreover, subcategory specificity cannot explain the asymmetries in exemplar usage because the target compounds were matched on this dimension.

The familiarity account offers reconciliation between the apparently contradictory exemplar and prototype theories. Both theories may accurately portray the judgment of compound categories at different stages of development; relatively unfamiliar compounds follow the exemplar account, whereas relatively familiar ones conform to the prototype theory. Using previous work on representational development (e.g., Klein & Loftus, 1993; Sherman, 1996) to extrapolate from our present data, our familiarity account would predict that sufficient experience with male elementary schoolteachers would lead to the emergence of a specialized abstraction. For example, a person who worked in education and had extensive contact with male elementary schoolteachers is unlikely to call on individual exemplars in order to judge the group as a whole. Although our data speak directly only to compounds, we would expect that a similar developmental pattern would be observed for subcategories that are not formed by conjunctions (e.g., sprinters as a subset of athletes). In essence, the familiarity account alludes to an element of continuity between compound categories, non–conjunctive subcategories, and simple categories (e.g., the engineers and club members in the work by Sherman, 1996).

ON CONSTITUENT COMPATIBILITY

This research goes well beyond a mere replication of Sherman's (1996) results at the level of the subordinate (compound) social category. For, in the case of compound categories, the compatibility of the constituents is an additional important factor that was not relevant in any of the earlier research conducted by Sherman, Klein, and their colleagues. Indeed, constituent compatibility has been hypothesized to play a central role in judgments of com-

pound categories (e.g., Hastie et al., 1990; Thagard, 1997), and provides competing hypotheses to those of the familiarity–based model.

The current results suggest that at least one important judgment mode—exemplar–based processing—is not determined by compound compatibility, per se. According to compatibility models, the features of compound categories are determined primarily by conceptual combination processes, in which the features of the constituents are somehow averaged to generate subcategory knowledge. Exemplar–based processes and other higher order reasoning processes are hypothesized to occur only when the constituent categories of the compound are incompatible, and resolution between them is not possible. Yet, in Experiment 3, in which the male teacher compound was perfectly compatible, exemplars were accessed as part of the judgment process. This points to the key role of compound familiarity rather than compatibility as a determinant of exemplar use.

We do not wish to suggest that constituent compatibility is an unimportant variable in judging compound categories. We do not know the extent to which incompatibility triggered complex reasoning processes in judging the male teachers in Experiments 1 and 2. The presence of exemplar–based processing does not rule out the possibility that other processes were also invoked.

We also do not know the extent to which compatibility may have influenced the use of conceptual combination processes in judgments of the female teachers. Indeed, we do not know that such processes did not contribute to judgments of the male teachers in Experiment 3. We do know that in the latter case, compatibility was not sufficient to preclude the use of exemplars. Compatibility by itself also could not explain the overall pattern of exemplar use across the three experiments.

Finally, it is important to note that compatibility is undoubtedly a critical variable in determining judgment processes for compounds that are so unfamiliar that no exemplars are known (e.g., Harvard–educated carpenters). This has been the nature of the compounds studied in most of the research examining the role of compatibility in judgment processes (e.g., Hastie et al., 1990; Kunda et al., 1990), and it has been argued that only such cases are appropriate for understanding the processes of concept combina-

tion (e.g., Medin & Rips, 2005; Rips, 1995). In these cases, there is no recourse to exemplars. All that is available is superordinate–based combinatorial strategies (in the case of compatible constituents) and complex reasoning and analogy processes (in the case of incompatible constituents) that also are based on general knowledge of the constituents.

Thus, the purpose of the present research is not to argue that compound compatibility is unimportant, but rather it is to examine the roles of familiarity and compatibility in the use of exemplar–based processing when exemplars are available. Though it may be true that the presence of known exemplars may "taint" pure conceptual combination process (e.g., Medin & Rips, 2005; Rips, 1995), the data presented here also demonstrate that such exemplars do play an important role when they exist. If we want to know how people come to understand compound categories that are actually encountered in the world, it will be necessary to consider the role of compound exemplars.

SELECTING AN INFERENCE STRATEGY FOR JUDGING COMPOUND CATEGORIES: A MODEL

The immediate aim of the present studies was not to pit exemplars and other bases for inference against each other, but rather to provide direct evidence for the use of exemplars and for the conditions that promote their use. However, these results can be integrated with previous research to put forward a more general model that describes which strategy will be used under what circumstances, when judging compound categories of a social or non–social nature. The model is described in some detail elsewhere, (Sherman & Groom, 2005) so we will only summarize it here.

Is the Compound Familiar Enough to Support Abstraction?. We argue that the first critical variable is familiarity with the compound itself (i.e., not only with the constituents). If the perceiver possesses specialized, abstract knowledge relevant to the judgment, resulting from sufficient experience with the compound category, then the decision process will stop here. Theoretically and empirically, simple deduction seems to be the easiest and most favored

judgment strategy. Familiarity with the compound can arise from information learned about specific members of the compound, or from information learned about the compound in general (e.g., being informed by your mother that male elementary schoolteachers are intelligent).

Are there Exemplars Available?. If the compound is not familiar enough to support abstraction, then perceivers will search memory for exemplars of the compound that could supply an inductive answer. This strategy was observed consistently across our three studies when judgments were required for male elementary schoolteachers, a relatively unfamiliar compound. All the participants in the current studies had access to relevant exemplars either from the experimental materials or from their own experience (the handful of participants in Experiment 1 who failed to recall an exemplar were thrown out).

Are the Constituent Categories Compatible?. There will be occasions when a perceiver has no relevant category exemplars stored in memory. For example, the cases of Harvard–educated carpenters and blind marathon runners used in the work by Kunda et al. (1990) are such categories (as are most of the compounds examined in the literature). In the event that no suitable exemplars are available at all, the perceiver must go beyond knowledge that directly pertains to the compound. In this case, judgments can proceed in one of two ways. First, the values of the constituent category abstractions (e.g., Harvard graduates and carpenters) on the judged feature dimension can be averaged, or combined according to some other, relatively simple and pre–existing rules (e.g., Anderson, 1981; Hampton, 1987; Smith, Osherson, Rips, & Keane, 1988). Second, the perceiver can make a creative attempt to synthesize the constituents into a coherent structure, drawing on complex cognitive processes such as causal reasoning, and on background information about the world. These are the types of processes described by Hastie et al. (1990) and Kunda et al. (1990).

It is here that compatibility–based dual–process models of conceptual combination (e.g., Hastie et al., 1990; Thagard, 1997) are likely to apply, distinguishing coherence–driven processing from incoherence–driven processing. When the constituents are relatively compatible, algorithmic procedures for conceptual combination would seem to have most relevance (e.g., Anderson, 1981;

Hampton, 1987; Smith et al., 1988). Conversely, Kunda et al.'s (1990) account of the construction of causal narratives as a route to judgment would apply most to judgments of incompatible, unfamiliar compounds (e.g., in deriving the features of a female mechanic or Republic social worker).

In short, we propose that extensional processes that are derived from knowledge of the compound itself will first be applied to judgments of the compound. If abstractions are available they will be used; if not, then exemplars will be used. Intensional processes based on category knowledge of the constituents alone will be applied primarily when judges have no familiarity with the compound itself. These are the conditions most often examined in the conceptual combination literature. In these cases, compounds based on compatible constituents will be judged via some combinatorial algorithm. In contrast, compounds based on incompatible constituents will be judged via complex reasoning processes.

CONCLUSION

In the terminology of the conceptual combination literature, our data support the notion that compound category representations are non–compositional. In other words, the features of compound categories (at least those that we actually encounter) are determined primarily through direct exposure to the compound itself (e.g., experiences with males who are not teachers or with teachers who are female would have little influence on judgments of male teachers). Regarding the direct use of exemplars in such judgments, one might question the prevalence of categories for which we have some exemplars, but not enough to derive abstract knowledge. However, we would argue that such processes are critically relevant for stereotype change processes whenever individuals join groups from which they have been excluded historically. In these situations, such "token" group members are endowed with considerable responsibility and power. Consider, for example, Black Americans in the Republican Party. Because such compounds do not include enough prominent members for people to have abstracted stereotypes about them, individual compound members are likely to have a great deal of personal in-

fluence on how public perceptions of that group develop. Thus, the personal political fortunes of Colin Powell and Condoleeza Rice are likely to have much greater implications for structuring people's impression of Black Republicans than do Dick Cheney's for White Republicans. Indeed, because voters have the power to determine whether further exemplars occur, the developmental nature of compound category impression formation may have profound social implications.

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, *11*, 211–227.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality & Social Psychology, 41*, 656–670.

Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27–58.

Gray, K. C., & Smith, E. E. (1995). The role of instance retrieval in understanding complex concepts. *Memory & Cognition*, *23*, 665–674.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer Jr, & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–68). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, *15*, 55–71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 12–32.

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society*, *28*, 242–247.

Klein, S. B., Babey, S. H., & Sherman, J. W. (1997). The functional independence of trait and behavioral self–knowledge: Methodological considerations and new empirical findings. *Social Cognition, 15*, 183–203.

Klein, S. B., & Loftus, J. (1993). The mental representation of trait and autobiographical knowledge about the self. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 5, pp. 1–49). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality & Social Psychology, 63*, 739–753.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14*, 551–577.

Medin, D. L., & Rips, L. J. (2005). Concepts and categories: meaning, memory, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 37–72). Cambridge England: Cambridge University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes, *Psychological Review, 84*, 231–259.

Rips, L. J. (1995). The current status of research on concept combination. *Mind and Language, 10*, 72–104.

Rosenthal, R. (1991). *Meta–analytic procedures for social research*. Newbury Park, CA: Sage.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality & Social Psychology, 70*, 1126–1141.

Sherman, J. W., & Groom, C. J. (2005). *Judging compound categories*. Unpublished Manuscript.

Sherman, J. W., & Klein, S. B. (1994). Development and representation of personality impressions. *Journal of Personality & Social Psychology, 67*, 972–983.

Sherman, J. W., Klein, S. B., Laskey, A., & Wyer, N. A. (1998). Intergroup bias in group judgment processes: The role of behavioral memories. *Journal of Experimental Social Psychology, 34*, 51–65.

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*, 337–361.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*, 485–527.

Smith, E. R. (1990). Content and process specificity in the effects of prior experiences. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 3, pp. 1–59). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

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

Thagard, P. (1997). Coherent and creative conceptual combinations. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 129–141). Washington, DC, USA: American Psychological Association.