Memory & Cognition
2002, 30 (4), 647-656

Dissociations between categorization and similarity judgments as a result of learning feature distributions

JEAN-PIERRE THIBAUT, MYRIAM DUPONT, and PATRICK ANSELME
University of Liège, Liège, Belgium

A dissociation between categorization and similarity was found by Rips (1989). In one experiment, Rips found that a stimulus halfway between a pizza and a quarter was categorized as a pizza but was rated as more similar to a quarter. Smith and Sloman (1994) discussed these results in terms of the role of necessary and characteristic features. In two experiments, participants had to learn to categorize novel artificial shapes composed of a nonsalient necessary feature combined with a salient characteristic feature. Participants categorized stimuli on the basis of a necessary feature, whereas their similarity judgments relied on characteristic features. The role of deep (essential) features in dissociations is considered. Results are discussed in terms of the differences between requirements of categorization and similarity judgments.

According to many authors, similarity is a central concept for models of categorization in the sense that categorization is grounded in similarity. Technically, an object X is categorized in Category A instead of Category B if its representation is more similar to the representation of Category A than to the representation of Category B (see Goldstone, 1994; Hampton, 1998; Komatsu, 1992; Murphy, 2002; Thibaut, 1997, for reviews of the relevant literature). This similarity-based view is one of the dominant theories of categorization.

Contrary to this view, Rips (1989; see also Rips & Collins, 1993) provided empirical evidence for a dissociation between categorization and similarity judgments. The experimental setup consisted of a comparison between categorization and similarity judgments of a target stimulus X with respect to two categories of stimuli, A and B. The rationale was that if the target was categorized in A more often than in B but judged more similar to B than to A, this result would demonstrate the dissociation. In one experiment, Rips read his participants a description of a target object described in terms of a value on a single dimension (e.g., the diameter). This value was chosen halfway between the largest dimensional value of a small category and the smallest value of a large category. To illustrate, a target 3-in. object was chosen to be halfway between participants' estimate of a U.S. quarter size (1 in.) and their size estimate of the smallest pizza (5 in.). The variance along the critical dimension is different in the two categories. The size of a quarter is fixed, whereas the size of pizzas is much more variable. In the categorization condition, participants were required to categorize the 3-in. target object in one of the two categories. In the similarity condition, participants were asked to rate the similarity of the target with respect to the two categories. It was shown that whereas most (63%) categorization participants categorized the target in the variable category (e.g., pizza), most similarity participants (69%) found the target to be more similar to the fixed category (e.g., U.S. quarters). This important result was taken as evidence that categorization is not based on similarity. In the case of categorization, most participants seemed to follow a rule.

Smith and Sloman (1994) tried to replicate Rips's (1989) results in two experiments. The instructions encouraged participants to use rule-based categorizations by pointing to the existence of a feature sufficient for categorization in one of the two categories (by "rule" we mean a set of properties—features—that, if they are satisfied by a stimulus, will lead to its categorization in a particular category). In their first experiment, they did not replicate the dissociation obtained by Rips. They hypothesized that this could be due to a procedural difference, in the sense that in Rips's experiment participants were instructed to talk aloud as they made their choices, whereas Smith and Sloman's participants were not asked to do so. Participants were asked to do so in Smith and Sloman's second experiment. In the first experimental condition of this experiment (the sparse condition), which was equivalent to Rips's experimental condition, choices in the similarity task (50% of choices in favor of the variable category) significantly differed from the results in the categorization task

Part of this research was presented at the Nineteenth Annual Meeting of the Cognitive Science Society, August 1997 in Stanford and at the Simcat Conference, November 1997, in Edinburgh. The authors thank Rob Goldstone and Philippe Schyns for discussions on this research and Gregory Murphy for his generous comments on this manuscript. Steven Sloman and two anonymous reviewers also contributed useful comments. Correspondence should be addressed to J.-P. Thibaut, Department of Psychology (Bat 32), 5 Boulevard du Rectorat, 4000 Liège, Belgium (e-mail: jpthibaut@ulg.ac.be).

(67%). However, they did not replicate one result obtained by Rips. In Rips, the similarity judgments were clearly in favor of the fixed category (69% of choices), but not in Smith and Sloman (50% in favor of the fixed category).

In a second experimental condition (the rich condition), Smith and Sloman (1994) added a characteristic feature of the fixed category to the sparse description of the target item (e.g., the characteristic feature "that is silver colored" was added to the original description "a circular object with an X-inch diameter"). The purpose of this condition was to contrast the *necessary-feature* hypothesis with the *characteristic-feature* hypothesis. According to the necessary-feature hypothesis, participants categorize objects solely on the basis of a necessary feature whenever one is available. The characteristic-feature hypothesis holds that people's categorizations are based on characteristic *and* necessary features when both kinds are available. In the rich condition, contrary to the sparse condition, participants categorized the target items in the fixed category (i.e., the quarter, 77%) and judged them more similar to this category (74%). According to the authors, these results seem to corroborate the characteristic-feature hypothesis. Indeed, in the rich condition, participants categorized the items more often in (and judged them as more similar to) the fixed category than in the sparse condition. This difference should not be expected if participants were relying exclusively on the necessary feature. In sum, Smith and Sloman obtained a small dissociation only in the sparse condition and under think-aloud instructions.

It is possible that many of the so-called characteristic features used by Smith and Sloman (1994) in the rich condition were not interpreted as more characteristic or less essential than the hypothesized necessary features; complementarily, it is also possible that the "necessary features" were not considered as more necessary than the characteristic features. This might be due to the way the authors chose exemplars of both classes of features. The nature, "essential" or "characteristic," of the features was not established independently but instead "by the authors' judgment" (p. 379). It is possible that the size in the pizza-quarter example was considered as highly implausible of the fixed category of quarters. It is also possible that the added "characteristic" feature of the object was considered as a very implausible property of the stimuli of the variable category (e.g., being silvered for pizzas). Complementarily, they could also have considered that "silvered" is as necessary of quarters as the size of quarters is, in the sense that a "nonsilvered" quarter would be considered as a false coin (made of a wrong metal). After all, quarters have both fixed size and material/color. Smith and Sloman noted that in the rich condition, when participants mentioned a characteristic feature in their spontaneous comments, they also categorized the stimulus in the fixed category. This result is compatible with the authors' characteristic-feature hypothesis but also with the idea that participants did not perceive the characteristic features as less necessary than the hypothesized necessary features.

Consider now the possibility that the hypothesized fixed-category-necessary features were not considered as necessary features. In many cases, the fixed categories used by Rips (1989) or Smith and Sloman (1994) were human activities defined by arbitrary rules or objects defined by arbitrary values. In these situations, participants might have considered that even though the defining feature is fixed by a rule (e.g., of the activity), in particular situations, it can be changed without modifying the nature of the activity (e.g., the official number of players—five—in a basketball game). These features are not central in the sense defined by Sloman, Love, and Ahn (1998), for whom "feature centrality is a function of the extent to which other features depend on it" (p. 191).

In other experiments, dissociations were obtained in the context of a distinction between deep and superficial features. The assumption was that categorization is influenced by "essential" (deep) features, whereas similarity is influenced by "superficial" (surface) features. In these studies, participants would use causal–explanatory theories about the world, and those theories would influence categorization. People tend to hold essentialist beliefs about natural kinds, and these categories are believed to have essential features that determine membership in the category. For example, features such as "parents are dogs" or "gave birth to a zebra" or "has a certain genetic structure" are supposed to be deep features, necessary for categorization. On the other hand, superficial features (e.g., animal color) are supposed to be characteristic features that should influence similarity judgments more than categorization. In one experiment, Rips (1989) provided evidence of dissociation between categorization and similarity judgments by manipulating these two types of features. Participants read stories about animals undergoing radical transformations. For example, a bird was transformed by a chemical accident into something that looks like an insect. In the categorization task, participants judged that the animal was more likely to belong to the bird category, whereas in the similarity task they judged the animal to be more similar to an insect. Thus, such an experiment seems to corroborate the necessary-feature hypothesis, since participants' categorizations followed deep features ("born from birds") rather than surface features. However, Estes and Hampton (2002) replicated Rips's transformation study only when they asked participants to perform both typicality and categorization judgments (within design), whereas in a between-participants design there was no dissociation. Moreover, the authors found that even when the dissociation was obtained in Experiment 1, it was true of only a minority of participants (in their data, less than 30%). Thus, again, the dissociation was obtained only in one context and only for a minority of people, which is not a very compelling argument in favor of the necessary-feature hypothesis (see also Hampton, 1995; Kalish, 1995; Malt, 1994).

However, these results should be interpreted cautiously. First, these descriptions refer to very uncommon and implausible events that, in turn, might lead to decreased participants' reliance on necessary features (e.g., a zebra transformed into a new animal that behaves like a horse). Second, as mentioned by Hampton (1995), it is not obvious that participants really know and follow the assumed

difference between the necessary feature (genotype) and the characteristic feature (phenotype). It is plausible to assume that most participants are not completely aware of the laws governing the biological world, and thus may ignore which features are necessary (see also Ahn & Dennis, 2001). If one cannot establish which properties a priori categorized as necessary (or characteristic) features by the experimenter were really interpreted and used as necessary (or characteristic) features by participants, one cannot decide which role is played by a property—necessary or characteristic—in a particular categorization.

In sum, Rips (1989) found a dissociation, and Smith and Sloman (1994) replicated it in their sparse condition under think-aloud instructions. Even in the latter case, the dissociation was quite small and, as stressed by Estes and Hampton (2002) in a similar context, it might have been limited to a small percentage of participants. In each case, the authors hypothesized that participants were engaged in some kind of analytic processing. The authors claimed that these results favor the characteristic-feature hypothesis for categorization. However, we have suggested that the status of features, characteristic or necessary, was questionable in these experiments.

The main purpose of our paper was to provide evidence that the dissociations really exist and to determine the context in which they appear. For that, we manipulated the role of necessary and characteristic features. This is important since many categorization models (such as prototype- and exemplar-based models) are similarity-based models (Hampton, 1995, 1998; Nosofsky, 1986; Rosch, 1978) and do not predict any dissociation between similarity and categorization.

Note that in Smith and Sloman's (1994) sparse description, where the dissociation was obtained, stimuli were described along only one dimension. The size in the pizza-quarter example had the role of a necessary feature in the fixed category and of a characteristic feature in the variable category. This situation is not optimal for contrasting the role of characteristic and necessary features in similarity rating and categorization. One purpose of the present experiments was to provide evidence of dissociations with multidimensional stimuli by manipulating the features constituting stimuli. We created novel artificial stimuli composed of features that appeared in an a priori fixed proportion of the stimuli: Stimuli were composed of features present in all the stimuli belonging to one category (necessary features) and of features present in a subset of stimuli belonging to one category (characteristic features). With stimuli perfectly controlled in terms of these constitutive features, one gets better control of the conditions eliciting dissociations, if any.

EXPERIMENT 1

We designed stimuli made of two features, each feature being chosen among three types: first, *necessary* features—that is, features that can be used as a rule for categorization because they were present in each stimulus of one category and absent in all the stimuli belonging to the second category. Second, *characteristic* features were present in a subset of each category's items. Third, we also introduced *neutral* features present in both categories—that is, features that have no diagnostic role. The characteristic features were designed to be salient, whereas necessary features were not salient in the sense that a careful analysis was needed in order to find them. Indeed, if the defining features were both salient and perfectly predictive and characteristic features were not salient and not perfectly predictive, there would be no reason to rely on characteristic features either for categorization or for similarity judgments. Another reason for contrasting non-salient-defining features with salient-characteristic features is that rule-based models and similarity-based models do not rely on these two classes of features in the same way (see predictions below).

We compared two experimental conditions that differed in terms of the association between the characteristic feature and the two categories. In the first condition (the *restricted* condition in what follows), the characteristic feature of Category A was absent in Category B and the characteristic feature of Category B was absent in Category A. Each characteristic feature was diagnostic because it was a perfect cue, though limited, for one category.

In the second condition (hereafter the *cross-category* condition), each characteristic feature was associated with one category in the same way as in the previous condition except that it was also associated with one stimulus of the other category. Thus, each characteristic feature was statistically diagnostic though it was not a perfect cue for categorization (see predictions below). This condition was introduced because with natural categories, and especially in experiments with artificial stimuli, when a (nondefining) feature is strongly associated with one category, it also appears in stimuli of contrastive categories.

The experiment was divided into a learning phase and a test phase. In the test phase, new stimuli were presented for classification. The purpose of the experiment was to compare similarity judgment and categorization obtained for transfer stimuli, called "incongruent"—that is, built with the characteristic feature of one category and the necessary feature of the other category. Dissociations should occur if the two judgments do not rely on the same type of feature.

Results obtained by Smith and Sloman (1994) and the characteristic-feature hypothesis predict that participants should not categorize solely on the basis of the necessary feature; some participants should categorize the incongruent stimuli according to the characteristic feature. More generally, similarity-based approaches to categorization also predict that characteristic features should influence categorization because they are salient (see models by Hampton, 1995; Medin & Schaffer, 1978; Nosofsky, 1986; Smith & Medin, 1981). According to these models, salient characteristic features should have a large impact on similarity judgments and categorization. Similarity-based models predict that if there is a difference between the restricted and the cross-category conditions, categorization should be more influenced by the characteristic features in the restricted condition. Thus, according to these mod-

els, no dissociation should be obtained, especially in the restricted condition. By contrast, rule-based models predict that categorization should be influenced only by the defining feature.

For similarity judgments on incongruent stimuli, similarity-based models predict that similarity judgments should be driven by characteristic features, whereas rule-based models have no strong prediction.

Dissociations should occur if participants use the nonsalient necessary feature for categorization and the characteristic salient feature for similarity judgment (or the reverse, which is unlikely). According to similarity-based models and Smith and Sloman's (1994) hypothesis, we should not expect such dissociations to occur. By contrast, rule-based models have no strong prediction about dissociations because they do not specify the role of characteristic features in similarity judgments. They would predict dissociations only if one assumes that even though people use a rule for categorization, they also register selected associations between a feature and a category, especially when the associated feature is salient.

Another purpose of the experiment was to show that dissociations are not the result of the use of deep causal features (e.g., a genetic cause) for categorization and superficial features for similarity judgments. If dissociations come through deep features, no dissociation should be obtained here given that our features had no deep causes but only statistical regularities.

Method

Participants. Twenty-two undergraduate students from the University of Liège volunteered for the experiment.

Materials. In the learning phase, two categories of 10 artificial stimuli were constructed. The stimuli were novel shapes that were composed of two parts, an upper part (the different F2 parts in Figure 1) and a lower part (the F1 parts in Figure 1). As mentioned above, there were two conditions, the restricted and the cross-category conditions. In the restricted condition, for 6 stimuli out of 10, the upper part, for Category A, has a mushroom shape slightly distorted over the 6 stimuli (F2a), and an angular shape for stimuli in Category B (F2b). These 6 stimuli were called "congruent" (panel A displays the 6 congruent stimuli of each category). The 4 remaining stimuli of the two categories, called the "neutral" stimuli, were constructed with four different upper parts (F2c,d,e,f). F2c,d,e,f are present in both categories and thus they cannot be used as cues for categorization (panel B displays the 4 neutral stimuli from the two categories). Each lower part is composed of four legs that are spatially grouped as one leg on the left and three legs on the right for Category A (1–3 legs; see Feature F1a in the stimuli), and two sets of two legs in Category B (2–2 legs; see F1b). The cross-category condition was constructed in the same way except that 1 of the 4 "neutral" stimuli from Category A and 1 neutral stimulus from Category B (SA10 and SB10, respectively) were replaced by 2 new stimuli. The first new stimulus was composed of the necessary (F1a) feature of Category A (i.e., 1–3 legs) and the characteristic feature of Category B (F2b); this stimulus was called "SA10′." The second stimulus was composed of one necessary feature from Category B (F1b) (2–2 legs) and one characteristic feature of Category A (F2a), a stimulus called "SB10′" (Figure 1C). We call these 2 contradictory stimuli "incongruent."

In sum, the cue validity of each characteristic feature (e.g., the probability that a stimulus is a member of Category A given that it has a mushroom shape) was 1 in the restricted condition. In the cross-category condition, the cue validity of the characteristic feature "mushroom shape" was .87 for Category A and .13 for Category B, whereas the reverse was true for the feature "angular."

In the test phase, 22 new stimuli, 11 per category, were constructed according to the same principles. For each category, there were 2 congruent, 4 neutral, and 5 incongruent stimuli.

Procedure. Each participant was randomly assigned to one of the two conditions—that is, 11 participants per condition. Participants were tested individually. They took 20–45 min to complete the task. The experiment was composed of two phases, a learning phase and a test phase.

In the learning phase, participants were told that they would have to learn to sort a set of stimuli into two categories. The initial stimulus was presented to the participant, who had to guess its category name (i.e., *moffo* and *quipi*). Feedback was provided about the accuracy of the answers. The next stimulus was presented in the same way, and so forth for the other stimuli. The order of presentation of the stimuli was random. Once the entire set had been presented to the participant, it was presented a second time, a third time, and so on, until the participant made no mistake during two successive presentations of the set of the stimuli.

In the test phase, participants were presented with the test stimuli one at a time. For each stimulus, the participant was asked to decide which of the two categories the stimulus belonged to and to choose the category the object was more similar to. Half of the participants performed the similarity task first, whereas the other half performed the categorization task first. This first task was followed by a rating task. Participants had to rate on a scale from 1 to 7 whether the test stimuli were likely to belong to Category A or to Category B. Similarly, they also had to rate whether the test stimuli were more similar to Category A or to Category B. The end of the scale corresponding to 1 referred to Category A and the end corresponding to 7 referred to Category B.

Results

First, we searched for dissociations between categorization and similarity judgments. As predicted, there was no dissociation for congruent stimuli. We analyzed the results obtained for the 10 incongruent test stimuli (i.e., 5 F1a + F2b stimuli and 5 F1b + F2a stimuli). We considered that a participant dissociated categorization and similarity judgment when he/she categorized 9 or 10 test stimuli in one category while estimating them more similar to the other category. Twelve participants (out of 22) produced such a dissociation. However, a comparison between the restricted and the cross-category conditions revealed that 10 participants (out of 11) dissociated in the restricted condition and only 2 in the cross-category condition. A Fisher exact test revealed that the proportion of dissociations obtained differed significantly in the two conditions ($p < .01$). To summarize, this analysis revealed that there were more dissociations between categorization and similarity judgments in the restricted condition than in the other condition.

In order to confirm these analyses, a two-way analysis of variance (ANOVA) (2×2) with category type (cross-category and restricted) as a between-participants variable and task (categorization and similarity) as a within-participants variable was performed on the ratings obtained for similarity judgments and for categorization. Dissociations are obtained for stimuli that get a small score for categorization and a high score for similarity or a small score for similarity and a high score for categorization.

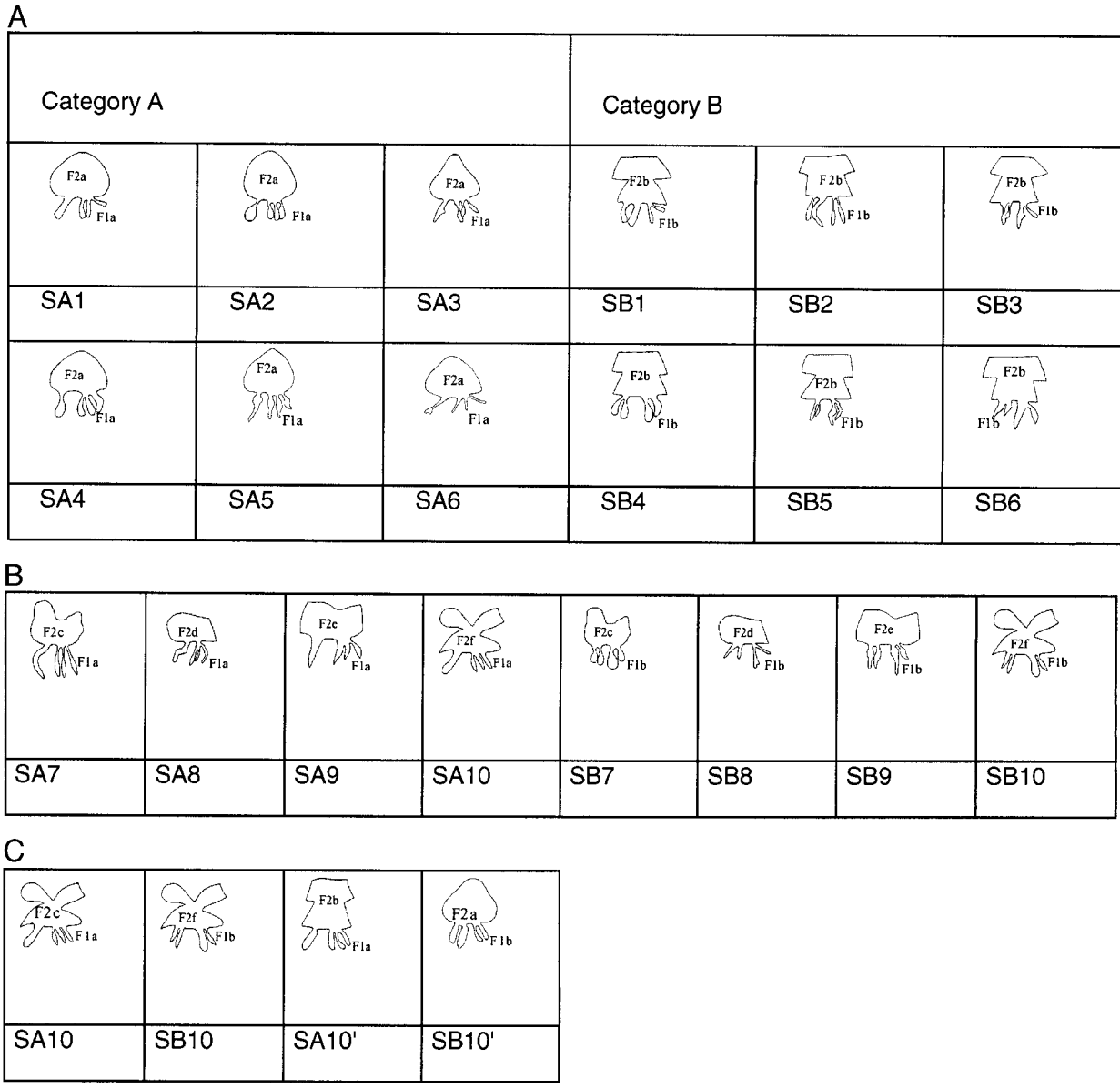

Figure 1. Panels A and B: The 20 stimuli from the restricted condition. Panel A: The 12 stimuli from Category A (SA1–SA6) and Category B (SB1–SB6) used in the learning phase. Panel B: The 8 neutral stimuli, SA7–SA10 for Category A, and SB7–SB10 for Category B. Panel C: In the cross-category condition, SA10 and SB10 were replaced by the "contradictory" stimuli SA10' and SB10'.

In order to perform a single analysis on the scores obtained for the test stimuli from both categories (1–3 and 2–2 test stimuli scores), we recoded the categorization and similarity judgment scores obtained for the 2–2 incongruent test stimuli (i.e., stimuli that had to be categorized in B when the participant followed the 2–2 rule). High scores in categorization, indicating that participants categorized the 2–2 stimuli in Category B, were transformed into small scores and vice versa. Small scores for similarity judgment, indicating that participants judged the 2–2 test stimuli as more similar to Category A, were transformed into high scores and vice versa. In other words, after transformation, a small score indicates that a participant categorized a stimulus in—or estimated it as more similar to—the category defined by the necessary feature, whereas a high score indicates that the stimulus has been categorized in—or estimated more similar to—the category defined by the characteristic feature. Results are shown in Figure 2. There was a significant effect of condition [$F(1,20) = 33.96$, $p < .0001$; results obtained in the restricted condition were higher than the equivalent result in the cross-category condition, 3.45 vs. 1.58, respectively], of task [$F(1,20) = 63.51$, $p < .0001$; $X = 3.9$ for similarity judgments vs. $X = 1.14$ for categorization], and a significant interaction [$F(1,20) = 25.77$, $p < .0001$; see Figure 2]. A posteriori test revealed a signifi-

cant difference between the restricted and cross-category conditions for similarity judgments (Tukey HSD, $p < .05$; 2.08 vs. 5.7 in the cross-category and the restricted conditions, respectively). There was no difference between these two conditions for categorization (1.08 vs. 1.2 for the cross-category and the restricted conditions, respectively). We computed a confidence interval on the similarity and the categorization scores in the two conditions at the level of $\alpha = .05$. The confidence intervals for the categorization scores were [0.81, 1.35] in the cross-category condition and [0.93, 1.47] in the restricted condition. For the similarity ratings, they were [1.06, 3.10] in the cross-category condition and [4.69, 6.72] in the restricted condition. In the restricted condition, the hypothesis that the mean would be beyond the value 4 (the intermediate value between 1 and 7 on the scale) was rejected for the categorization scores whereas it was accepted for the similarity scores.

In order to specify the lack of influence of characteristic features on categorization, we compared the categorization ratings obtained for the congruent and the neutral stimuli in both conditions. If the characteristic feature influenced categorization, we should get smaller ratings for congruent than for neutral stimuli. In both conditions, there was no significant difference between the two types of stimuli.

DISCUSSION

The analysis of the confidence interval confirms the results obtained for the ratings. The dissociation between similarity and categorization appeared, for the majority of participants, in the restricted condition, whereas no dissociation was obtained in the cross-category condition. The two conditions differed in terms of the similarity scores but not in terms of the categorization scores. In both conditions, categorization was based on the rule, whereas similarity judgments were based on the characteristic features in the restricted condition and on the rule in the cross-category condition. Our results also show that it is possible to obtain dissociations even when there is no deep, causal feature (vs. surface features) involved in the experimental design. The data did not confirm the characteristic-feature hypothesis since, in both conditions, characteristic features did not influence categorization.

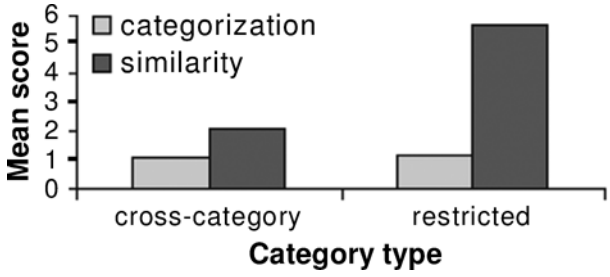

Figure 2. Interaction task × condition. Note the dissociation between categorization and similarity judgments in the restricted condition.

In the cross-category condition, even similarity judgments were not influenced by characteristic features. This suggests that participants' similarity estimation was driven by defining features only, by contrast with what was predicted by the above theories. However, one should interpret the cross-category results cautiously. One possibility is that, because of the incongruent stimuli, participants did not notice the association between each characteristic feature and one of the two categories. Given that the characteristic feature was salient, it was, most likely, the first feature tested by participants. Suppose that this first rule worked for a couple of trials; then, one contradictory item appeared, refocusing the participant's attention to other potential relevant features. For the rest of the experiment, even if participants noticed the presence of each characteristic feature, they would not pay attention to the category with which it was associated. The interspersed neutral features (6 stimuli) also obscured the relationship between each characteristic feature and its category. In sum, if this were true, the absence of influence of characteristic features on the similarity judgments in the cross-category condition would not be the result of a rigid rule-based behavior (driven by the presence of contradictory items) but the result of a misperception of the association between one characteristic feature and a particular category.

EXPERIMENT 2

In Experiment 2, a new condition was created in which, before learning, participants were asked to free sort the set of stimuli into two categories. A free-sorting task should give them an overview of the distribution of each characteristic feature in the stimuli and, because of their salience, participants should use them to free sort the stimuli, and later use them as features relevant for categorization at the beginning of the learning phase. If in the first experiment participants did not dissociate in the cross-category condition because they did not notice the association, they should dissociate in the present experiment, as in the restricted condition. On the other hand, if they had noticed the existence of the association in the cross-category condition of the first experiment and used necessary features for categorization and similarity judgment because of the presence of the contradictory items in the learning phase, again, we should obtain no dissociation between similarity and categorization. The restricted condition was also run in the same way as a control condition.

Method

Participants. Twenty-two undergraduate students from the University of Liège participated as volunteers in the experiment.

Material. The stimuli used in the first experiment were also used here. The two conditions compared in the present experiment were defined in the same way as in the first experiment: In the restricted condition, 6 congruent stimuli and 4 neutral stimuli and, in the cross-category condition, 6 congruent stimuli, 3 neutral stimuli, and 1 in-

congruent stimulus. The 22 test stimuli, 11 per category, used in the first experiment, were also used in the present experiment.

Procedure. Each participant was randomly assigned to one of the two conditions, for 11 participants per condition. Again, participants were tested individually. The experiment was composed of three phases: a free-sorting task, a learning phase, and a test phase. In the free-sorting task, participants were presented with the 20 stimuli designed for the learning phase and asked to sort the stimuli into two categories, the only constraint being that each category should comprise 10 stimuli. Then, participants moved to the learning phase and the test phase, which were the same as in Experiment 1.

Results and Discussion

Free-sorting task. Two participants found the correct rule for categorization (1–3 vs. 2–2). Eighteen participants used the upper part as a distinctive cue (mushroom vs. angular shape). They categorized the two members of a neutral pair (e.g., SA7 and SB7) in the same category, and the members of another neutral pair in the second category. Since in the cross-category condition there were only three neutral pairs, participants put 1 stimulus in each category in order to have the same number of stimuli in each category as required by the experimenter. The last 2 participants mentioned that they categorized according to the shape of the legs: their length, their thickness, or both.

Test phase. The main purpose of the present experiment was to replicate the results obtained in the restricted condition in the first experiment and to generalize these results to the cross-category condition when participants first analyzed stimuli in the free-sorting task. The number of dissociations obtained in the two conditions for the 10 incongruent test stimuli (i.e., 5 F1a + F2b stimuli and 5 F1b + F2a stimuli) was compared. A participant was categorized as a dissociator when he/she categorized 9 or 10 incongruent test stimuli in one of the two categories (A or B) while judging them more similar to the other category (B or A). Nine participants (out of 11) produced such a dissociation in the cross-category condition. The same number of participants were dissociators in the restricted condition. The proportion of dissociations obtained in the cross-category condition was also compared to the corresponding proportion (i.e., 2 dissociators) obtained in this condition in the first experiment. A Fisher exact test revealed that the two proportions differed significantly ($p < .01$). Thus, the results revealed that, with a free-sorting task added before the learning task, most of the participants dissociated categorization and similarity judgments in the cross-category condition.

As in the first experiment, a two-way ANOVA (2×2) with category type (cross-category and restricted) as a between-participants variable and task (categorization and similarity) as a within-participants variable was performed on the ratings obtained for similarity judgments and for categorization. Dissociations were obtained for stimuli with a small score for categorization and a high score for similarity or a small score for similarity and a high score for categorization. As in the first experiment, we recoded the categorization and similarity scores obtained for the 2–2 incongruent test stimuli (i.e., stimuli that had to be categorized in B when the participant followed the 2–2 rule). That is, after transformation, a low score indicates that a participant categorized a stimulus in—or judged it as more similar to—the category defined by the necessary feature, whereas a high score indicates that the stimulus has been categorized in—or judged more similar to—the category defined by the characteristic feature. There was a significant effect of task [$F(1,20) = 108.28, p < .0001$; $X = 1.34$ for categorization vs. $X = 5.37$ for similarity judgments]. There was no significant effect of category type [$F(1,20) = 0.0, p > .1$], and, by contrast with Experiment 1, no significant interaction [$F(1,20) = 0.01, p > .1$]. As in Experiment 1, we computed a confidence interval on the similarity and the categorization scores in the two conditions at the level of $\alpha = .05$. The confidence intervals were [1.05, 1.60] for the categorization scores in the cross-category condition and [1.08, 1.63] in the restricted condition. For the similarity ratings, they were [4.33, 6.36] in the cross-category condition and [4.36, 6.40] in the restricted condition. These results show a dissociation in both conditions. In fact, the hypothesis that the mean is beyond value 4 (the intermediate value between 1 and 7 on the scale) was rejected for the categorization scores whereas it was confirmed for the similarity scores.

As in Experiment 1, we compared the categorization ratings obtained for the congruent and the neutral stimuli in both conditions in order to assess the role of characteristic features on categorization. In both conditions, there was no significant difference between the two types of stimuli. For similarity ratings, note that the difference between congruent and neutral stimuli was significant in both conditions [$F(1,10) = 73.8, p < .001$, and $F(1,10) = 53.83, p < .001$ for the cross-category condition and the restricted condition, respectively], where neutral stimuli got higher scores than congruent stimuli, indicating that they were judged as less similar to their category members than were congruent stimuli. This is consistent with the idea that characteristic features influenced similarity judgments.

The present experiment replicates the results obtained in the restricted condition in the first experiment. The experiment also revealed a dissociation in the cross-category condition, contrary to what happened in Experiment 1. This suggests that the absence of dissociation in the first experiment was due to participants failing to notice the association between each characteristic feature and a category. Note that the result is potentially surprising because free-sorting tasks are notorious for leading people to one-dimensional sortings (Medin, Watenmaker, & Hampson, 1987; Spalding & Murphy, 1996). Thus, the prediction that participants would notice that the dimension they used in the free-sorting task was associated with other dimensions was not obvious because, after all, if they noticed the association, they should use it in their sorting.

GENERAL DISCUSSION

Our main purpose was to study the role of characteristic and defining features in similarity judgments and categorization, particularly in the case of dissociations between

two tasks (Rips, 1989; Smith & Sloman, 1994). Our results showed that, when the frequency of association between characteristic and necessary features and each category was controlled, no participant categorized incongruent stimuli in the category associated with the characteristic feature, thereby suggesting that participants did not use the characteristic feature in their categorizations. By contrast, similarity judgments were strongly influenced by characteristic features. We will discuss these results in the light of the characteristic-feature hypothesis and similarity-based models and in the context of the aforementioned association between dissociations and deep features.

In their experiments, Smith and Sloman (1994) observed a small dissociation in their sparse condition under think-aloud instructions and no dissociation in their rich condition. By contrast, we obtained a dissociation in a situation that was at least as rich as their rich condition, showing that dissociations can be obtained with multidimensional stimuli and for a majority of participants. As mentioned in the introduction, one possible reason why the authors did not get a sharper dissociation was that participants did not interpret the features as characteristic or necessary, contrary to what they were supposed to be. Note also that our data do not confirm the characteristic-feature hypothesis proposed by the authors.

Second, a number of authors have suggested that the dissociations should be obtained when surface and deep features define the experimental situation (Kroska & Goldstone, 1996; Rips, 1989; see, however, Estes & Hampton, 2002; Hampton, 1998). The introduction stressed that the status of the features, either deep or surface, was often debatable and this made the interpretation of the results less straightforward. Recently, Ahn and Dennis (2001) conducted experiments in which the status of the features— deep versus superficial features (in the authors' terms, cause features vs. effect features)—was controlled and dissociations were obtained. According to the causal status hypothesis (Ahn, 1998), features that cause other features receive more weight in categorization judgment and in the framework of psychological essentialism (Medin & Ortony, 1989). The authors hypothesized that categorization would be more influenced by deep features than similarity judgments. In their paradigm, a target situation is defined by three features: A that causes B that causes C. Consider now two situations, one sharing a single deep feature (Feature A) with the target, the other situation sharing two surface features with the target (Features B and C). If the deeper features are weighted more heavily in categorization judgments than in similarity judgments, the first situation should be selected more frequently in the categorization task than in the similarity judgment task. The results confirmed this view. Though deeper features influenced similarity judgments (56% of the participants judged the situation sharing one deep feature with the target more similar to the target), categorization was more influenced by these deep features (75% of the participants categorized the one-deep-feature-shared situation with the target). Ahn and Dennis argued that feature weighting is determined by the causal status of features in both similarity and categorization judgments and that cause features matter more than effect features in categorization. They concluded that there is a tendency to give more weight to surface features than people would in categorization judgment and, in that sense, categorization does not depend on similarity (see also Medin & Ortony, 1989; Rips, 1989). Our results show that this statement is restrictive in that dissociations can be obtained in the absence of deep features.

Why Participants Dissociate Similarity Judgments and Categorization

Estes and Hampton (2002) suggested that dissociations were obtained in situations that favor some kind of "reflective processing." This account is also compatible with Smith and Sloman's (1994) result that dissociations were obtained only when participants were engaged in a think-aloud situation. As suggested by Estes and Hampton, these situations may promote a kind of contrastive processing, in the sense that participants who are asked to make two (or three) successive judgments may seek ways to differentiate them and thus be induced to dissociate their ratings. This does not explain why people always dissociate in the same direction—that is, why similarity judgments are more influenced by surface features and categorization is more influenced by deep features. Also, the contrastive hypothesis does not explain the dissociation obtained by Ahn and Dennis (2001) using a between-participants design. Our design was a within-participants design and thus our dissociations could be the result of the contrasting hypothesis. However, recall that half of the participants were asked to perform the similarity judgments first, whereas the other half had to begin with categorization. Consider participants who dissociated. If we compare the categorization results of those who started with categorization with the similarity judgment results of those who started with similarity judgments, we have a perfect between-participants dissociation because all the categorization-first participants categorized according to the defining feature and all the similarity judgment-first participants used the characteristic feature in their estimation of similarity.

We believe that dissociations will appear each time the contrasted categories have features that people interpret as defining (even if these features are not really defining), whereas they interpret other features—the characteristic features—as associated with one particular category without being defining of this category. Because of the learning phase, this was the case with our characteristic and defining features. In terms of the control of features, the paradigm used by Ahn and Dennis (2001) is related to ours. They achieved the difference between deep and surface features through the manipulation of the causal connections between features, whereas in our situation the defining and characteristic features got their status during the learning phase.

In order to explain the orientation of dissociations (similarity goes with surface or characteristic features and categorization with deep or necessary features), one has to assume that people weight features differently depend-

ing on the task. Why would this be the case? According to Ahn and Dennis (2001), a mechanism underlying their dissociation is that when deep features are defined as those that cause surface features, deep features are weighted more in categorization judgments than in similarity judgments. The reason for this difference in weighting is, according to the authors, that people believe that the surface features of the members of a category share the same underlying cause. On the other hand, unless the goal of the similarity judgment is specified, the relative weighting one feature gets depends on and varies with context and task. In our case, categorization was not associated with deep features. Thus, we have to rephrase this statement in more general terms: Features that are thought to be perfectly predictive of a category with respect to the other possible categories will receive more weight in categorization than other features. Features used in similarity judgments depend more on the context. Since the number of possible contexts is infinite (Murphy & Medin, 1985), there is a large number of features that can be referred to in similarity judgments.

The importance of defining features in the case of categorization is illustrated by our restricted condition. In this condition, both types of features were perfect cues for categorization, the only difference being that the defining feature was associated with all the stimuli in one category, the characteristic feature being associated only with a subset of one category stimuli. In this case, categorization was not influenced by characteristic features (see also the absence of difference between the ratings obtained for the congruent and the neutral stimuli).

One also has to explain why participants did not choose the same features in both tasks. We believe that when there is no particular goal specified for the similarity judgment, participants will use features that encompass the largest part of the compared entities. In our situation, the characteristic feature was the most salient one because it constituted virtually the totality of the stimulus. In a similarity judgment for which features must be chosen, it may seem natural to rely on features that are the most salient and to neglect other features, unless there is something in the task that requires going beyond the most obvious features available. Thus, if the defining feature had been more salient than the characteristic feature, no dissociation would have been obtained (in Experiment 2, most participants in the free-sorting task spontaneously sorted stimuli according to the salient characteristic feature, whereas a small minority used the defining feature). Similarly, in Ahn and Dennis (2001), participants weighted the two characteristic features more in similarity judgments than in categorization. Interestingly, they showed that when the similarity instructions were "Consider all of the information available. Which option (A or B) is the target more like?" instead of the standard instructions, there was no dissociation between similarity and categorization. Relatedly, Goldstone (1994) argued that the task demand may force participants to interpret "similar" as "visually similar." However, he did not clearly explain why this may have been the case. We would say that unless participants are pushed to look beyond the immediate situation, they have no reason to search for less obvious features. By contrast, instructions such as "consider all the information available" may motivate them to search for features that are not directly available in the stimuli (see Medin, Goldstone, & Gentner, 1993; Thibaut & Schyns, 1995, for discussion of the notion of setting a feature space for categorization and similarity judgments).

Dissociations and Similarity Models of Categorization

As mentioned in the introduction, many models of categorization rely on the idea that categorization is a matter of similarity computation. This idea is central in prototype models that assume that categorization is a matter of comparison between the target and prototypes stored in memory. The idea is also central for exemplar-based models (Kruschke, 1992; Nosofsky, 1986). These similarity-based models do not account for our results. For example, consider Nosofsky's (1986) generalized context model or Kruschke's (1992) ALCOVE model. According to these models, participants learning the categories in our experiments will shift their attention to the leg dimension because it perfectly predicts the categories. This feature will be heavily weighted in the classification decision—and because the classification decision is based on similarity, the similarity between an instance and a category will also be strongly affected by this dimension, which is in conflict with our data. On the other hand, if participants start first with the salient characteristic feature, in the restricted condition the characteristic feature always points to the same category; even if the defining feature gets more attention as learning proceeds, there is no reason why this feature would not affect categorization, whereas it remains central for similarity judgments. Clearly, characteristic features did not affect categorization since there was no difference between neutral and congruent items.

Recently, a number of categorization models have attempted to classify instances using rules and similarity to previously experienced exemplars (see Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). Hybrid models should explain why participants learned nothing about the association between the characteristic feature and its category in the cross-category condition, or why the contradictory stimuli encountered during learning did not influence the test phase. Indeed, these items presented during learning should be weighted as specific exemplars belonging to one category and influence categorization of the contradictory items at test. One should obtain a difference between the cross-category condition and the restricted condition for the contradictory items, which was not confirmed by the data.

In general, in similarity models, the weight of each feature is a matter of its diagnosticity and salience in the category with which it is associated (Thibaut & Schyns, 1995). To capture the dissociations studied here, one should add a mechanism that emphasizes the role of features that are interpreted as necessary by people (even

though these features are not very salient) and that decrease the importance of salient characteristic features in the case of categorization, whereas similarity judgments seem to be determined more by the salience of the features involved in the comparison.

REFERENCES

Ahn, W. K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, **69**, 135-178.

Ahn, W. K., & Dennis, M. J. (2001) Dissociation between categorization and similarity judgement: Differential effect of causal status on feature weights. In U. Hahn, & M. Ramscar, (Eds), *Similarity and categorization* (pp. 87-107). New York: Oxford University Press.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

Estes, Z., & Hampton, J. A. (2002). *Similarity and essentialism in category judgments of natural kinds*. Manuscript submitted for publication.

Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, **52**, 125-157.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory & Language*, **34**, 686-708.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, **65**, 137-165.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, **23**, 335-353.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, **112**, 500-526.

Kroska, A., & Goldstone, R. L. (1996). Dissociations in the similarity and the categorization of emotions. *Cognition & Emotion*, **10**, 27-46.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology*, **27**, 41-70.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254-278.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.

Medin, D. L., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

Medin, D. L., Watenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, **19**, 242-279.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology*, **122**, 468-486.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition*, **22**, 377-386.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 525-538.

Thibaut, J. P. (1997). Similarité et catégorisation. *Année Psychologique*, **97**, 701-736.

Thibaut, J. P., & Schyns, P. G. (1995). The development of feature spaces for similarity and categorization. *Psychologica Belgica*, **35**, 167-185.

(Manuscript received September 27, 2000;
revision accepted for publication January 9, 2002.)