Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 54 (2006) 407–424

Journal of Memory and Language

www.elsevier.com/locate/jml

The importance of being coherent: Category coherence, cross-classification, and reasoning ☆

Andrea L. Patalano [a,*], Seth Chin-Parker [b], Brian H. Ross [c]

[a] *Department of Psychology, Wesleyan University, 207 High St., Middletown, CT 06459-0408, USA*
[b] *Denison University, USA*
[c] *University of Illinois at Urbana-Champaign, USA*

Received 25 July 2005; revision received 20 October 2005
Available online 10 January 2006

Abstract

Category-based inference is crucial for using past experiences to make sense of new ones. One challenge to inference of this kind is that most entities in the world belong to multiple categories (e.g., a jogger, a professor, and a vegetarian). We tested the hypothesis that the *degree of coherence* of a category—the degree to which category features go together in light of prior knowledge—influences the extent to which one category will be used over another in property inference. The first two experiments demonstrate that when multiple social categories are available, high coherence categories are selected and used as the basis of inference more often than less coherent ones. The second two experiments provide evidence that ease of category-based explanation of properties is a viable account for coherence differences. We conclude that degree of coherence meaningfully applies to natural social categories, and is an important influence on category use in reasoning.

Keywords: Categories; Reasoning; Inference; Coherence

A critical function of categorization is inference (Heit, 2000; Smith & Medin, 1981). Once an entity has been identified as a member of a known category, a wealth of category knowledge can be used to reason about that entity. A challenge for category-based inference is that most entities belong to multiple categories. Cross-classification is important to consider because people rarely incorporate information from more than a few categories (Heit & Rubinstein, 1994; Murphy & Ross, 1999; Ross & Murphy, 1999), so they must somehow solve the problem of selecting among available categories. The goal of this paper is to integrate cross-classification issues with recent category coherence research towards understanding how people make novel property inferences about cross-classified entities.

Cross-classification

Past research has identified three influences on category preference when more than one category is available. First, people are more inclined to use the

☆ This work was supported by National Science Foundation Grant SBR 97-20304. We thank Nick Haslam and Gregory Murphy for comments on an earlier draft of this paper. Thanks also to Melissa Paulson and Jane Erickson for their assistance in data collection.
* Corresponding author. Fax: +1 860 685 2761.
E-mail address: apatalano@wesleyan.edu (A.L. Patalano).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2005.10.005

category with the greatest relevance to the property in question (Heit & Rubinstein, 1994; Kalish & Gelman, 1992; Murphy & Ross, 1999; Ross & Murphy, 1999). Second, inferences are more often made from categories with increased mental activation relative to others (Macrae, Bodenhausen, & Milne, 1995; Sinclair & Kunda, 1999; Smith, Fazio, & Cejka, 1996). And third, greater emphasis is placed on the most distinctive category available (Nelson & Klutas, 2000; Nelson & Miller, 1995; van Rijswijk & Ellemers, 2002), where distinctiveness refers to the relative number of members of one category over another in a particular situation or in the population at large.

One of Nelson and Miller's experiments (1995; Exp. 3) influenced the design of our own experiments, and serves as a good example of the kind of situation under discussion. Problems were used such as: "80% of dog owners prefer non-fiction to fiction. 80% of skydivers prefer fiction to non-fiction. Bob is a dog owner and a skydiver. Which is he more likely to prefer, non-fiction or fiction?" Each problem paired one high distinctiveness (e.g., sky diver) and one low distinctiveness (e.g., dog owner) category, and used properties previously unrelated to either category. Participants chose the more distinctive category 69% of the time.

One striking absence in this research is the exploration of factors independent of context or of a specific property in question. This leaves open the important question of whether some structural properties of categories might promote inference more than others. In the next section, we motivate the study of one such structural factor, namely, *category coherence*.

Category coherence

Category coherence refers to the extent to which category features go together in light of prior theoretical, causal, and teleological knowledge (Medin, 1989; Murphy & Medin, 1985; see Murphy, 2002, for a review) rather than being just incidentally co-occurring. "Lives in water, eats fish, has many offspring, is small" describes a more coherent category than "lives in water, eat wheat, has a flat end, is used for stabbing bugs" (Murphy & Wisniewski, 1989). It is well documented that most natural categories are at least somewhat coherent (Ahn, 1998; Keil, 1989; Malt & Smith, 1984; Sloman, Love, & Ahn, 1998), and that coherence of novel categories influences ease of learning and use (Heit & Bott, 2000; Kaplan & Murphy, 2000; Lin & Murphy, 2001; Murphy & Allopenna, 1994; Pazzani, 1991; Rehder & Ross, 2001; Spalding & Murphy, 1996; Wattenmaker, Dewey, Murphy, & Medin, 1986; Wisniewski, 1995).

The relations that make features "go together" can be causal (Ahn, 1998; Rehder & Hastie, 2001, 2004), spatial or temporal (Lin & Murphy, 2001), abstract themes (Erickson, Chin-Parker, & Ross, 2005; Rehder & Ross, 2001), or goals (Barsalou, 1983, 1985). While all facilitate learning, there is some evidence that a *common cause* structure, one in which a few causal features give rise to many effect features, results in especially high category coherence (Ahn, 1998; Ahn & Kim, 2000; Ahn, Kim, Lassaline, & Dennis, 2000; but see Rehder & Hastie, 2004; Sloman et al., 1998). This structure is consistent with *psychological essentialism* (Medin, 1989; Medin & Ortony, 1989), the finding that people believe that entities in the world have deep underlying features that are enduring and unchangeable even though their surface features might change (Atran, 1990; Hirschfeld, 1994, 1996; Keil, 1989; Rothbart & Taylor, 1992; Yulill, 1992). We will focus on the common cause structure in this paper; it is what will be meant by coherence unless otherwise noted.

The coherence of a category has been shown to influence category-based inference. Using experimental methods, Rehder and colleagues (Rehder & Burnett, 2005; Rehder & Hastie, 2004; see also Lassaline, 1996) found that people are more likely to transfer a property from a category to a new member when the category is causally coherent rather than incoherent. Haslam, Rothschild, and Ernst (2000) conducted a factor analysis on the Likert-scale ratings of 40 social categories (including jobs, racial and ethnic groups, hobbies, religious groups, etc.) on nine dimensions. An emergent "entitativity" factor approximated coherence, and was associated with three scale items targeting common-cause structure: inherence (the presence of deep underlying features giving rise to surface ones), uniformity (the similarity of category members), and informativeness (the inference potential of a category).

While the coherence of natural categories can be assessed through entitativity scale ratings and elicitations of mental representations (e.g., listings of deep underlying category features), most inference experiments have used artificial categories. Because artificial categories are typically created to be maximally coherent versus maximally incoherent, it has been difficult to assess the effects of everyday variations in coherence on inference.

Current research

The present research was guided by two major goals. The first goal was to consider the extent to which category coherence influences category use in reasoning from multiple categories. In particular, when high and low coherence categories are placed in direct competition with one another, are higher coherence categories favored over less coherent ones? The second goal was to begin to explore cognitive processes underlying differential use of high versus low coherence categories, looking at how people use their knowledge about categories to explain their property inductions. After a pretest of category materials, we conducted four experiments to address these goals. In addition, in the context of the pretest and Experiment 1, we reassessed the influence of distinctiveness suggested by Nelson and Miller (1995), in light of our observation of a possible confound between distinctiveness and coherence.

Pretesting of categories

In using natural categories, we must specify how the degree of coherence for particular categories is established. For example, consider the categories of "ministers" and "county clerks." Our intuition is that, consistent with a common cause notion of coherence, ministers are associated with deep underlying traits such as belief in God, compassion for others, and satisfaction in attending to the spiritual needs of a community. These features give rise to surface behaviors such as being on a first name basis with people in the community, living in a parsonage, and working on the weekends. While the latter features are common among ministers, they derive from deeper properties of individuals who become ministers. This contrasts with a category such as county clerk for which underlying unifying properties might not be available. Such intuitions might be insightful, but a more specified process for determining the coherence of the categories used in the study is necessary.

Our starting point for identifying the coherence of various social categories was a laboratory database of job and hobby categories that had been previously rated by undergraduates on similarity ("How similar are two category members to one another?" on a 1–7 Likert-scale), and frequency ("How many people in the United States per 1000 would you estimate to be a member of this category?). We took advantage of these similarity ratings as an initial basis for identifying high versus low coherence categories. Our decision was based on the premise that similarity judgments are, at least in part, a function of coherence. All else being equal, the more coherent a category, the higher is its rated similarity (Lassaline, 1996; Rehder & Hastie, 2004).

Six pairs of categories differing on similarity, but matched on frequency, were chosen as possible high versus low coherence category pairs. Six pairs matched on similarity, but differing on frequency, were chosen as high versus low distinctiveness pairs. Categories were randomly chosen from the database with the constraint that job and hobby categories be used in the same proportion, and that paired categories be applicable to a single individual. See Appendix A for the resulting 12 category pairs.

In the following pretests, to provide more direct evidence for coherence differences, we collected entitativity scale ratings (Haslam et al., 2000) and deep feature listings (Ahn, 1998; Rehder & Burnett, 2005; Rehder & Hastie, 2004) for the coherence categories. Note that while we focused here on one set of categories in order to provide broad evidence for coherence, later studies incorporated other categories as well. In these pretests, we also obtained frequency, similarity, and entitativity ratings for the Nelson and Miller (1995) categories to assess the possibility of a confound between distinctiveness and coherence.

Method

Participants

At the University of Illinois, 17 undergraduates participated in an entitativity task, and a different 24 undergraduates participated in a feature listing task, in exchange for monetary compensation. All participants were tested in groups of 1–5 in 1-h sessions. The tasks were presented in booklets; each took about 15 min to complete. They were preceded by an unrelated categorization task.

Entitativity procedure

On the first two pages, similarity ratings and frequency estimates (as previously collected for our own database) were collected for Nelson and Miller categories. On the remaining three pages, ratings on the three entitativity scales (uniformity, informativeness, and inherence) were collected for both our coherence categories and the Nelson and Miller categories. See Appendix B for exact wordings of Haslam et al.'s (2000) entitativity scales. For each scale, the definition appeared at the top of a page followed by all categories. Categories were presented in a different random order for each of the three scales.

Feature listing procedure

For coherence categories only, participants were instructed to "Describe in as much detail as possible what the members of each category are like in a *deep* sense. In other words, what are the important and essential characteristics of someone who is in each category." This was the only definition of deep features given to participants. Booklets contained subsets of three of the six category pairs. Within a booklet, categories appeared in random order with three categories per page.

Results

Entitativity results

High and low coherence category ratings on frequency and similarity measures (from the earlier established database) and on the three entitativity measures are shown in Table 1. Similar ratings for the Nelson and Miller categories appear in Table 2. For the coherence categories, the categories we had identified as high coherence items were given higher ratings than low coherence ones on all entitativity measures (uniformity: $t(16) = 7.75$, $p < .001$; informativeness: $t(16) = 7.76$, $p < .001$; and inherence: $t(16) = 4.47$, $p < .001$); this pattern held for all pairs of categories. This supports the claim that the category pairs differ in coherence. For the Nelson and Miller categories, on both similarity and entitativity measures, the three high distinctiveness items were given higher ratings than the three low distinctiveness ones, with no overlap in distributions. This provides evidence of a confound between coherence and distinctiveness in the materials used by Nelson and Miller (1995).

A Pearson correlation matrix (see Table 3) shows pairwise correlations between entitativity measures, similarity ratings, and frequency estimates (the latter two coming from our preexisting database) for our categories. Strong positive correlations (range = .81 to .97) were found between pairs of entitativity measures and between each entitativity measure and similarity ratings ($p < .001$ for these comparisons). Correlations with frequency estimates ranged from −.18 to −.35 (negative because entitativity and similarity estimates increase as frequency *decreases*) but were not statistically significant ($p > .100$). The results replicated those of Haslam et al. (2000) in showing that category uniformity, informativeness, and inherence co-vary across categories, consistent with the notion that coherence is a meaningful construct represented by these three scale items.

Feature listing results

An average of 3.9 features were generated for each coherence category. As shown in the last column of

Table 1
Frequency, similarity, entitativity, and feature listing results for pretest job and hobby categories

	Frequency	Similarity	Entitativity scale			Features
			Uniformity	Informativeness	Inherence	
High coherence						
Soldier	32	4.3	7.1	6.7	5.5	4.9
Feminist supporter	55	4.5	6.3	6.6	5.5	3.9
Minister	12	4.9	7.0	7.2	6.2	5.8
Pro wrestler	3	5.4	4.8	6.8	5.1	4.7
Yacht club member	16	4.7	6.5	6.8	4.9	3.8
Rare-sculpture collector	6	4.6	6.1	5.7	4.5	3.6
M	21	4.7	6.3	6.6	5.3	4.5
Low coherence						
Matchbook collector	17	2.9	5.0	4.4	4.4	2.7
Waiter	73	2.3	2.4	2.5	2.1	4.2
Rubber-stamp collector	11	3.1	5.7	4.8	4.3	3.2
Badminton player	14	2.4	3.3	2.8	2.9	3.0
County clerk	10	2.8	4.8	3.8	3.1	3.5
Limousine driver	13	3.1	3.5	3.1	2.0	4.6
M	23	2.8	4.1	3.6	3.1	3.5

Note: High and low coherence category pairs (items matched on frequency but contrasting on coherence) are in the same position in the high and low coherence lists respectively (e.g., soldier and matchbook collector).

Table 2
Frequency, similarity, and entitativity results for Nelson and Miller (1995) categories

	Frequency	Similarity	Entitativity scale		
			Uniformity	Informativeness	Inherence
High distinctiveness					
Snake owner	39	4.4	5.5	4.9	4.7
Sky diver	82	5.2	6.1	6.0	4.6
Six siblings	35	4.0	4.4	4.5	4.1
Low distinctiveness					
Dog owner	397	2.4	3.0	2.8	1.9
Tennis player	196	3.5	4.2	3.4	3.5
One sibling	389	3.4	2.9	2.8	1.9

Table 3
Coherence (Uniformity, informativeness, and inherence), similarity, and frequency correlations for pretest job and hobby categories

	Uniformity	Informativeness	Inherence	Similarity
Uniformity				
Informativeness	.97*			
Inherence	.94*	.95*		
Similarity	.87*	.94*	.81*	
Frequency	−.35	−.20	−.18	−.25

Note. Similarity and frequency ratings were obtained from earlier participants (in database).
* $p < .01$.

Table 1, a greater number of features were generated for high as compared with the low coherence categories ($t(23) = 3.36$, $p = .003$); this pattern held for 4/6 pairs of categories. Given that the task was to generate only deep features, these results alone provide reasonable evidence of a difference. However, because it was desirable to obtain independent confirmation that participants did in fact generate deep features, three independent coders (two of the three were blind to our hypothesis) rated each feature on a 1 (surface) to 7 (deep) Likert-scale. Pairs of coders were in agreement, defined as being within one point of each other, on 76–79% of all problems.[1] The average rating of the three coders was used to determine depth. The mean rating for high coherence features ($M = 3.6$, $SE = .03$) was reliably higher than that for low coherence ones ($M = 3.2$, $SE = .03$), $t(23) = 2.84$ $p < .009$; the patterns of results were the same for each coder separately as well.

Summary of pretesting

The pretest provided evidence that we have a set of six category pairs that vary within each pair on coherence—based on similarity, entitativity, and deep feature listings—but not on distinctiveness. It also allowed us to confirm that the Nelson and Miller (1995) categories confounded distinctiveness and coherence, motivating a consideration of each factor separately as it relates to category use in inference. Recall that we also previously established a set of six category pairs that vary on distinctiveness but not on coherence. While it might be that coherence and distinctiveness are correlated in natural categories, it is now possible to tease apart the effects of these two characteristics to better understand their influences on inference when multiple categories are available.

[1] One coder's responses were an average of 1 point lower than the others, but with the same variance. This coder's responses were linearly transformed by adding 1 point to them. This made it easier to compare values across coders but did not influence any statistical results.

Experiment 1: Basic inference

With materials from the pretest in hand, we were now able to develop a set of inference problems involving entities belonging to multiple categories. Specifically, problems similar in structure to those used by Nelson and Miller (1995) were developed. These problems asked participants to make property inferences about individuals belonging to two categories, with the difficulty being that the categories provided conflicting information. For six coherence problems, high and low coherence categories were put in conflict. For six distinctiveness problems, high and low distinctiveness categories were in conflict. We hypothesized that participants would be more inclined to reason from high as compared with low coherence categories. We had no strong prediction in regard to the distinctiveness categories.

Method

Participants

Twenty-six undergraduates at the University of Illinois participated in exchange for monetary compensation. Participants were tested in groups of 1–9 in 20-min sessions.

Categories

The sets of coherence category pairs and distinctiveness category pairs, drawn from our initial database as previously described, were used for all problems. See Appendix A for both sets.

Materials

Twelve problems of the following format were created.

Imagine that the following information is true:
80% of feminist supporters prefer Coca-Cola to Pepsi.
80% of waiters prefer Pepsi to Coca-Cola.
Chris is a feminist supporter and a waiter.
What beverage does Chris prefer, Coca-Cola or Pepsi?_

(Not confident at all) 1-2-3-4-5-6-7 (Extremely confident).

The first premise always stated that 80% of members of the category in question had a certain preference, while the second stated that 80% of members of a different category had a complementary preference. A fictitious individual (with a gender-neutral name not related to either category) was introduced as a member of both categories. The task was to infer this individual's preference and to assign a confidence rating.

Properties were assigned with care to ensure that paired properties matched as closely as possible on desirability and base rate. Properties also had no obvious relationships with their assigned categories. And they were stated so as to be mutually exclusive and exhaustive in the context of the problem.

Materials were presented in a booklet with four problems per page. The coherence problems were presented first in a single random order, followed by the distinctiveness problems. There were two booklet versions. Version A used the exact materials shown in Appendix A, while Version B reversed the two categories within each problem. The latter served not only to change the order of the categories, but also to pair them with different properties (e.g., in the earlier example, the category "waiter" would come first and would be paired with a preference for Coca-Cola).

Procedure

Participants were instructed to work on the problems, in order, at their own pace. The entire task took ≈15 min to complete.

Results

Coherence problem results

The main dependent measure was the percentage of times the more coherent category was chosen. On average, the higher coherence category was chosen the majority of the time ($M = 66\%$, $SE = 5.1$; $t(25) = 3.58$, $p = .001$). Individually, 5 out of 6 problems were consistent with the overall results. There were no differences in confidence ratings for high ($M = 3.7$, $SE = .32$) versus low ($M = 3.3$, $SE = .24$) coherence responses ($t(21) = 0.53$, $p = .600$); four participants were excluded from the statistical analysis of confidence because all of their responses fell into one (high coherence) response category.[2]

Distinctiveness problem results

The main dependent measure was the percentage of times the more distinctive category was chosen. Surprisingly, the *less* distinctive category was chosen the majority of the time ($M = 65\%$, $SE = 4.1$, $t(25) = 3.72$, $p = .001$) counter to the predictions of the distinctiveness hypothesis of Nelson and Miller (1995). All individual problem results were consistent with the overall results. There was no difference in confidence ratings for high ($M = 3.6$, $SE = 0.29$) versus low ($M = 3.5$, $SE = 0.31$) distinctiveness responses ($t(22) = 0.64$, $p = .530$); three participants were excluded from the statistical analysis because all of their responses fell into one (low distinctiveness) response category.[3]

Discussion

The results provide evidence that coherence is related to category use for inference, with high coherence categories having greater influence than low coherence ones. The results offer no evidence that distinctiveness influences category use in the way suggested by Nelson and Miller (1995). Counter to their hypothesis, low distinctiveness categories were selected more often than high distinctiveness ones. One possible explanation is that distinctiveness provides clues as to which property has the highest base rate, and that this information is used when more category-specific strategies are not available.

One limitation of the first experiment was that participants were given all inference-relevant information. In many everyday contexts, however, people have to gather this information themselves, and must decide which information to seek out. The resource costs (e.g., time and effort) of getting information in situations might lead people to obtain only information that they already perceive to be the most relevant to the task at hand. If the usefulness of a category is assessed prior to obtaining further information about an individual's membership in the category and the category's properties, then this is a stage in reasoning in which heuristics such as use of coherence are very likely to come into play. Experiment 2 expanded on the findings of Experiment 1 by exploring the role of category coherence in selecting information in the course of reasoning from multiple categories.

A second limitation of the first experiment was that because the problems used forced a choice of one category or the other, it was not possible for participants to identify multiple categories as important or to integrate information from multiple categories if desired.

[2] In other, unpublished studies, we found a similar pattern of results (using various probabilities in the materials) when participants had to write the *probability* that the individual would have a given property.

[3] These results are not dependent on using known social categories. The same results were obtained with both serially and sequentially learned artificial categories.

An even stronger test of the coherence hypothesis would be one in which the option was available to incorporate information from both high and low coherence categories. If individuals were to still focus on high coherence categories, we would have even stronger evidence for the use of coherence in inference among multiple categories. The information selection task used in Experiment 2 allowed for the selection of and integration of information from multiple categories if desired.

Experiment 2: Information selection

In Experiment 2, we developed problems in which participants could select categories about which they desired inference-related information. For example, one could choose the category "waiter" and find out that "20% of waiters prefer Coca-Cola to Pepsi." For each problem, there were two high and two low coherence categories; participants could choose as few as one category or as many as all four. The dependent measure was the number of high versus low coherence categories about which information was requested. This procedure was modeled after Murphy and Ross (1999; see also Spellman, Lopez, and Smith, 1999).

The problems used in this experiment were modified from those in Experiment 1 in the following ways. First, the target individual for each problem was identified as being a member of four categories rather than two. Second, the relationship between each category and the property in question was not initially presented. Rather, participants were instructed to choose categories—as many as they wished—that they deemed relevant to the judgment at hand in order to reveal the property information about the categories. Third, rather than giving a binary response as in Experiment 1, participants were asked to assess the likelihood that the hypothetical individual had the property in question (though this was not the focus of the study).

We expected that high coherence categories would be chosen earlier and more often than low coherence ones. If high coherence categories are perceived to be more informative for inference, people should prefer to gain property information about these categories when such information is not initially present.

Method

Participants

Sixteen undergraduates at the University of Illinois participated in exchange for monetary compensation. Participants were tested in groups of 1–5 in 1-h sessions. This task followed an unrelated categorization task.

Categories

Categories used were the coherence items from Experiment 1, supplemented with additional database categories. Within each problem, categories varied in coherence; the highest to lowest coherence categories in each problem (as approximated by database similarity ratings) were labeled high, medium–high, medium–low, and low (see Appendix C for Experiment 2 materials). On average, high coherence categories had a similarity rating of 4.7 (on the 1–7 scale), medium–high coherence categories were 3.9, medium–low were 3.2, and low were 2.4. Categories within a problem were roughly matched on frequency and were combined such that an individual could plausibly be a member of all four categories.

Materials

Ten problems were created following the format of Fig. 1. For each problem, four categories were presented on the left side of the display, and the percentage of members of each category having a given property was presented on the right side of the display, though the latter were hidden when the problem first appeared. A hypothetical individual was described as being a member of all four categories and the participant was asked to

Consider the following information about various social categories:

SKY DIVER	▷	<10% prefer Coca-Cola to Pepsi>
MATCHBOOK COLLECTOR	▷	<45% prefer Coca-Cola to Pepsi>
MUSEUM GUARD	▷	<55% prefer Coca-Cola to Pepsi>
PROFESSIONAL WRESTLER	▷	<15% prefer Coca-Cola to Pepsi>

Based on this information, please make the following judgment:

John is a member of each of these categories. What is the likelihood that he prefers Coca-Cola to Pepsi?

Fig. 1. Problem presentation format used in Experiment 2. Information in angle brackets, < >, was not presented for its corresponding category unless participant clicked the mouse on that category.

assess the likelihood that the individual had some property in question.

The properties were a subset of those used in Experiment 1 (see Appendix C). Property probabilities were divisible by 5 and were chosen from the following ranges in approximately equal number: 10–30%, 40–60%, or 70–90%. Within a problem, the probabilities for the two highest coherence categories were drawn from one range, and the probabilities for the two lowest coherence categories were drawn from a different range.

Two versions of the materials (Versions A and B) were created. Version B was the same as the Version A with the following exceptions: category sets and corresponding properties were paired differently, feature components were reversed (e.g., "prefers Coca-Cola to Pepsi" became "prefers Pepsi to Coca-Cola"), the probabilities assigned to the two highest and two lowest coherence categories within each problem were swapped, and category positions (first, second, third, or fourth in list) within a problem were reordered (see Appendix C for details). Across versions of the materials, each coherence level appeared in each category position in the display an equal number of times.

Procedure

Problems were presented one at a time on a Macintosh computer using HyperCard 2.4 software. Each problem initially appeared without feature probability information. Participants were instructed that: "You may request information about as few or as many categories as you wish, with the constraint that you must choose at least one category. It is important, however, that you only request information that you believe to be highly relevant to your judgment." Information about each category could be selected by clicking on the arrow to the right of each category. After requesting all desired information, participants entered a probability judgment between 0 and 100% before going on to the next problem.

Participants were given a practice problem and had an opportunity to ask questions before beginning the target problems. Assignment of participants to Versions A and B of the materials was alternated. Participants worked at their own pace, taking ≈10 min to complete all problems.

Results

Four participants (two using each version of the materials) chose all four categories in the same order for every problem, suggesting no attempt to modulate their responses in regard to category relevance as instructed. These participants were not considered in the analyses. For the remaining 12 participants, no differences were found in the pattern of results for material Versions A and B, so the results are reported together.

Table 4
Number (percentage) of categories chosen at each coherence level in each selection position in Experiment 2

Selection position	Coherence level			
	High	Med high	Med low	Low
1st (100% of problems)	40 (33%)	35 (29%)	29 (24%)	16 (13%)
2nd (83% of problems)	27 (27%)	36 (36%)	20 (20%)	17 (17%)
3rd (55% of problems)	11 (16%)	15 (23%)	19 (29%)	21 (32%)
4th (31% of problems)	10 (27%)	05 (14%)	06 (16%)	16 (43%)

Note. All problems are represented in the 1st selection position because participants were instructed to select at least one category per problem. Number of problems represented at each subsequent selection position decreases because participants did not always select a second, third, or fourth category for a problem.

An average of 2.7 categories were selected per problem. Table 4 shows how many categories of each coherence level were chosen in each temporal selection position (first, second, etc.), aggregating over all participants and problems. For example, the first category selected per problem was a high coherence category 40 times (33%), a medium–high coherence 35 times (29%), a medium–low coherence 29 times (24%) and a low coherence 16 times (13%). A one-way within-subjects ANOVA on just this first selection position, in which all categories were available for selection, revealed a statistically-significant linear trend ($F(1,11) = 5.41$, $MSE = 2.33$, $p < .050$). Participants were reliably more likely to choose categories of higher coherence for the first position. Because later responses were dependent of the first one (e.g., once a high coherence category was chosen in the first position, it could not be chosen again in the second position), similar analyses could not be conducted for each position.

To consider both the number of times each coherence level was chosen and in what temporal position, points were assigned to each coherence level: 4 points each time it was chosen first, 3 points for second, 2 points for third, and 1 point for fourth. This allowed us to establish a total number of points for each coherence level for each participant. The means of these totals from highest coherence to lowest were, respectively: 23 ($SE = 1.7$), 24 ($SE = 1.4$), 18 ($SE = 1.4$), and 14 ($SE = 1.7$). A one-way within-subjects ANOVA on these totals revealed a statistically reliable linear trend ($F(1,11) = 21.49$, $MSE = 29.50$, $p < .001$); there was a preference for the two higher coherence categories over the two lower ones. This overall pattern of results held for 10 out of 12 participants.

While we had no strong prediction regarding probability judgments, we were nonetheless interested in the extent to which judgments were based on the probabilities associated with the higher coherence categories. For each of the 34 problems in which exactly one higher and one lower coherence category were selected, the distance between the two category probabilities was divided into equal thirds. Each judgment response was then coded as being in the third closer to the higher coherence category, in the middle third, or in the third closer to the lower coherence category. The number of responses falling into each third was 11 (32%), 13 (38%), and 10 (29%) problems, respectively. In other words, once categories had been selected, coherence no longer played a role in judgment.

We also pursued the possibility that rather than the most coherent categories being given the most weight, the earliest selected categories were given the most weight. To address this possibility, an analysis similar to the previously described was again performed. This time, each judgment was coded based on its proximity to the first versus second selected category. The number of responses falling into each third was 15 (44%), 13 (38%), and 6 (18%), respectively. While not statistically reliable with so few observations, the pattern suggests a tendency to place greater judgment weight on the first category selected.

Discussion

When given the opportunity to choose inference-relevant information, people showed a greater tendency to select information related to high coherence as compared with low coherence categories. High coherence information was selected earlier and more often than low coherence information. This finding replicates the general phenomenon established in Experiment 1, that category coherence guides category use in inference, and extends it to situations in which the opportunity to use information from both high and low coherence categories is available.

When exactly one higher and one lower coherence category were selected, there was no indication that categories of higher coherence were given more weight in computing probability judgments. Rather, responses were typically weighted towards the probability associated with the first category selected. This finding is consistent with earlier results to the extent that determination of category relevance is what guides order of category selection in the first place.

The results reveal a clear ability for people to differentiate the two highest from the two lowest coherence categories. The results do not reflect any finer grained discrimination, particularly between very-high and high coherence categories. Possible explanations for these results are that people are unable to make such fine-grained distinctions, that such distinctions are not considered necessary for this reasoning task, or that similarity imperfectly reflects coherence. Further work is needed to address this issue more fully.

In the situations considered thus far, the properties in question were chosen to be previously unassociated with the corresponding categories. And the presented information gave no clear reason for favoring one category over others based on prior knowledge. The best response would have seemed to be 50% in Experiment 1, and the average of each set of four probabilities in Experiment 2. Yet the properties and probabilities associated with the more coherent categories were used the majority of the time.

Experiments 3a and b: Single-category explanations

One possible cognitive explanation for the results of the first two experiments is that individuals engage in explanation-based reasoning about new properties. Sloman (1994; see also Heit and Rubinstein, 1994) found that people were more inclined to transfer a property from one category to another when they could generate a single coherent explanation for its presence in both categories. Sloman (1994) offered the example that "Many ex-cons are hired as bodyguards. Therefore many war veterans are hired as bodyguards" is more convincing than "Many ex-cons are unemployed. Therefore many war veterans are unemployed." The first argument can be covered by the common explanation that veterans and ex-cons have a past that would lead them to be physically tough, but the second cannot be easily covered by a single explanation.

People might also be more likely to transfer a property from a category to a member to the extent that they can generate a strong explanation as to why the property might be common among the category members. For example, willingness to transfer "prefers Coca-Cola" to a particular sky diver might increase to the extent one generates the explanation that: "Skydivers might prefer Coca-Cola because they are driven by a need for heightened sensation and, containing a large amount of caffeine, Coca-Cola provides this sensation." Transfer would be less likely when all available explanations are weak or when none come to mind.

How would explanation account for differences in use of high versus low coherent categories in inference? Forming an explanation involves generating a strong causal link from one or more known features of a category to the newly learned property. As shown in the pretest, high coherence categories offer more deep features on which to build new links and to integrate novel features into a rich theory of the category. As a result, high coherence categories might be more likely than low coherence ones to support the generation of convincing explanations.

In Experiment 3a, participants were given information about a hypothetical preference of members of a social category (e.g., that most soldiers prefer gin to whiskey), and were asked to generate the most plausible explanation for this preference. For half of the problems given to each participant, the categories were high in coherence; for the other half they were low in coherence. Categories were the same as those used in Experiment 1 that varied on coherence but were matched on frequency. After completing the problems, participants were asked to go back and rate each generated explanation for plausibility. We hypothesized that high coherence category explanations would be rated as more plausible than low coherence ones.

Because participants in Experiment 3a could have been influenced by having written the explanations themselves, in Experiment 3b, a separate group of participants was asked to rate the plausibility of each explanation generated by the earlier participants. More specifically, each participant in Experiment 3b was given a booklet from 3a to evaluate. We expected independent raters to again rate high coherence explanations as more plausible than low coherence ones.

The goal of these two experiments was to assess the extent to which category coherence predicts the quality of the explanations that individuals generate when they are asked to explain why members of a category possess a novel property. These experiments speak to the plausibility of an explanation-based account of differences in the use of high versus low coherence categories for inference. These studies are also interesting in their own right in that they allow us to replicate coherence differences in a related area of category-based reasoning, namely, the use of categories in explanation formation.

Note that neither of these studies explicitly involved cross-classified entities. If explanations influence reasoning about cross-classified entities, we should first see differences in the quality of explanations generated for high versus low coherence categories individually. If such differences exist, we would then have good reason to consider which categories people use as the source of explanation when multiple categories are available.

Method

Participants

Eight undergraduates at the University of Illinois generated and rated their explanations (Experiment 3a), and twenty-four undergraduates rated the explanations of the others (Experiment 3b) in exchange for introductory psychology course credit.

Materials

Booklets consisted of 12 problems, using the same coherence category pairs as Experiment 1. The problems were of the following format:

> Approximately, half of all people in the United States prefer vacationing in Bermuda over vacationing in the Bahamas. Among professional wrestlers, however, there is a strong preference for Bermuda over the Bahamas. Please generate the most plausible explanation you can think of as to why this might be the case.

The properties used were the same as those used in Experiment 1 for both coherence and distinctiveness problems. Two versions (Versions A and B) of the booklets were made, with different random orders of problems and different pairings of categories and properties in each version. Each property was paired with a high coherence category in one version and a low coherence category in the other version.

Procedure

In Experiment 3a, participants were tested in a single group in a 20-min session. They were given the booklet of problems and asked to work on it at their own pace. On the last page of the booklet, instructions asked participants to go back through the problems in order and to rate each generated explanation for plausibility on a scale of 1 (Highly implausible) to 7 (Highly plausible).

In Experiment 3b, participants were tested in groups of eight in 20-min sessions. The explanations from the eight booklets from Experiment 3a were typed into eight new booklets (so that the new participants would see typed rather than hand-written explanations), and the plausibility ratings from the Experiment 3a participants were omitted. Each of the eight booklets was given to three participants. Participants were asked to read the explanations and to rate each on a scale from 1 (Highly implausible) to 7 (Highly plausible).

Results

In Experiment 3a, the two booklet versions showed the same pattern of results and so the data were collapsed. The mean plausibility rating for high coherence categories was 3.8 ($SE = .31$) versus 3.1 ($SE = .35$) for low coherence ones, $t(7) = 2.44$, $p = .040$. A sample response to the high-coherence professional-wrestler problem was "Wrestlers are more daring and want to go to dangerous, risky areas such as Bermuda [over the Bahamas]" (rating = 4). A response by the same participant to a low-coherence rubber-stamp collector problem (where members "have a strong preference for tulips over roses") was "Rubber stamp collectors are passive and prefer lighter and softer colors, such as tulips [over roses]" (rating = 2).

In Experiment 3b, the two booklet versions again showed the same pattern of results and the data were collapsed. The high coherence explanations ($M = 4.0$, $SE = 0.14$) were again rated as more plausible than the low coherence ones ($M = 3.4$, $SE = .16$), $t(23) = 4.01$,

$p < .001$. In other words, both groups of participants rated the explanations as more plausible for high as compared with low coherence categories.

Discussion

The purpose of Experiments 3a and 3b was to investigate the extent to which category coherence predicts quality of category-based explanations. The results support the hypothesis that people generate better explanations for high coherence as compared with low coherence categories. While this effect might appear somewhat small, it should be considered in light of the following points. First, the only information available to participants was category membership, and participants were essentially forced to use this information to generate a response for each problem. Thus any category-coherence differences in willingness to generate an explanation could not be observed. Second, though related, participants were given unlimited time in which to generate responses. So the results do not consider any relative effort that may have gone into generating plausible responses for high versus low coherence categories. Third, any reliable difference in effect size is likely to be important in situations in which multiple categories are available as sources of explanation. As long as one category is deemed a better source of explanation, it may be more likely to be used to explain behavior in the context of multiple competing categories.

The last point leads us to the purpose of Experiment 4. One limitation of Experiment 3a and 3b was that it considered explanation only in the context of one category at a time. An important extension of this work is to consider whether or not category-based explanations are more likely to be generated from high as compared with low coherence categories when multiple competing categories are available. Such a finding would further strengthen the possibility of a link between category-based explanation and category use in inference from multiple categories.

Experiment 4

The materials used in this experiment were similar to those of Experiment 3a except that each problem described people who were members of two categories (one high and one low coherence category) rather than one. Participants were asked to generate three different explanations for the stated preference. At the end of the task, they were asked to go back and circle the most plausible explanation (from among the three) for each problem. We hypothesized that explanations would make reference to high coherence categories more often than to low coherence categories, especially among the "most plausible" explanations.

Method

Participants

Eighteen undergraduates at the University of Illinois participated in exchange for introductory psychology course credit.

Stimuli

The 12 coherence categories used in Experiment 1 were paired here to create six problems; six more problems were created using categories from the laboratory database (see Appendix D for all materials). Members of category pairs differed in similarity (4.6 versus 2.9 on average for high versus low coherence category sets, respectively) but were equated on distinctiveness (30/1000 people for both high and low coherence category sets). As in Experiment 1, high and low coherence category sets were matched on the ratio of job to hobby categories.

Materials

Booklets consisted of 12 problems, each problem containing one high and one low coherence category. The problems were of the following format:

> Approximately, half of all people in the United States prefer fiction over non-fiction. Among people who happen to be both weekend badminton players and professional wrestlers, however, there is a strong preference for fiction over non-fiction. Please list three separate plausible explanations as to why this might be the case.

The properties used in each problem (e.g., preferring Bermuda versus the Bahamas) were again chosen to have no prior association with the problem categories. One version of the booklet was created, with problems presented in a single random order.

Procedure

Participants were tested in groups of 6 in 30-min sessions. They were given the booklet of problems and asked to work on it at their own pace. On the last page of the booklet, instructions asked participants to go back through the problems and to circle the most plausible explanation for each one.

Results

Coding

Two students (one undergraduate and one graduate), unaware of the experimental hypothesis, were paid to code the data. For each problem, they were asked to decide whether the explanation made reference to only the first presented category, to only the second presented category, to both categories, or to neither category. The experimenter then recoded the results of the coders as follows: *High-coherence* (explanation makes reference only to the high coherence category), *Low-coherence* (explanation makes reference only to the low coherence category), *Both* (explanation makes reference to both categories), or *Neither* (explanation makes reference to neither category). Coders were instructed to only count a category as being mentioned if the participant ''made direct reference to the category.'' This could occur if the participant used the category itself in the explanation (e.g., ''professional wrestlers like danger...'') or if direct reference was made to a clear property of the category (e.g., people who fight one another in their jobs must like danger...'').

For each explanation, the responses of the two coders were combined by assigning 0.5 points to the category selected by each coder. Thus, if both coders chose the same code, the code for that explanation would receive a combined points value of 1.0, otherwise the two chosen codes would each receive 0.5 points. This was done after ensuring that the inter-rater agreement was high—raters were in agreement for 97% (627 out of 648) of the responses. The pattern of results would not have changed if either one or the other of the coder's responses, rather than both of them, had been used.

For each participant, we then computed the percentage of points given to each code. We did this separately for all explanations (*All-Explanations* analysis), and for the most plausible explanations only (one per problem, as identified by participants; *Plausible-Only* analysis).

Analyses

The results of the All-Explanations analysis will be described first. As illustrated in Table 5, for nearly half of the explanations (49%), participants made explicit reference to neither category, and gave an explanation for why anyone might prefer one over the other (e.g., ''it is warmer in the Bahamas''). These results are consistent with literature suggesting that people frequently ignore base rates in reasoning (e.g., Tversky & Kahneman, 1974). The results also likely reflect the inherent difficulty in generating explanations for previously unrelated properties, especially when there is little incentive to do so. In the remainder of cases, participants used one category or the other the majority of the time (39% of all explanations), and only rarely made reference to both categories (12% of all explanations). Of central importance to the present investigation was the relative use of the high versus low coherence category in the large subset of cases in which only one category was selected. In these cases, we found that participants relied on the high coherence categories (22% of all explanations; 56% for this subset of the data) reliably more often than the low coherence categories (17% of all explanations; 44% for this subset of the data; $t(17) = 2.58$, $p = .020$).

The results for the Plausible-Only analysis followed the same pattern but showed even greater reliance on high as compared with low coherence categories. Again, as illustrated in Table 5, approximately half of the time (53% of all most-plausible explanations), participants made reference to neither category. In the remainder of cases, participants used one category or the other the majority of the time (33% of all most-plausible explanations), and considerably less often made reference to both categories (14% of all most-plausible explanations). We were again interested in the relative use of high versus low coherence categories in the large subset of situations in which only one category was used, and found that participants relied on the high coherence categories (22% of all most-plausible explanations; 67% for this subset of the data) twice as often as the low coherence categories (11% of all most-plausible explanations; 33% for this subset of the data; $t(17) = 2.62$, $p = .020$). In other words, as with the results with all data, high coherence categories were used more often than low coherence ones in generating explanations.

Table 5
Coherence levels of categories used in explanations in Experiment 4

	High coherence	Low coherence	Both	Neither
All explanations	22% (3.0)	17% (3.3)	12% (3.1)	49% (7.5)
Plausible Only	22% (4.0)	11% (2.6)	14% (4.7)	53% (7.5)

Note: Standard errors are in parentheses.

Discussion

The purpose of this experiment was to assess the extent to which category coherence influences category use in explanation when multiple categories are available. The results are consistent with the hypothesis that high coherence categories are used more often than low coherence ones in generating novel explanations. In fact, when considering only the most plausible explanations, high coherence categories were mentioned twice as often as low coherence ones. These results build on Experiments 3a and 3b in which it was found that the more plausible explanations were generated for high coherence as compared with low coherence categories. The studies taken together suggest that the structures of high coherence categories are more conducive to the generation of plausible explanations for novel properties.

General discussion

Summary of results

The primary purpose of the experiments was to better understand the role of category coherence in reasoning about cross-classified entities. In pretests, we identified social categories that varied in similarity, one marker for coherence, and then provided confirmation that these categories also differed on other measures associated with coherence including entitativity (Haslam et al., 2000) and the presence of deep features (Ahn, 1998; Keil, 1989). We also found that distinctiveness and coherence were confounded in a study by Nelson and Miller (1995), suggesting coherence as an alternative explanation for their results.

Two experiments explored our focal interest, the influence of category coherence on category-based inference in the context of multiple social categories. In Experiment 1, when forced to choose between a high and a low coherence category in making an inference about a hypothetical individual belonging to both categories, participants used the high coherence category the majority of the time. In addition, in contrast to the findings of Nelson and Miller (1995), when high and low frequency categories were contrasted here, it was the high frequency (low distinctiveness) that promoted category use. In Experiment 2, four categories were available and the dependent measure was the coherence of the categories about which property information was requested. Again, higher coherence categories played a greater role in inference; in this case, higher coherence categories were selected earlier and more often than lower coherence ones.

Two experiments then considered one possible account for the coherence differences, namely, that people try to explain the presence of novel properties in terms of known category features and that this is easier to do for high coherence categories. This argument is consistent with the definition of coherence as one in which deep underlying features give rise to surface features. In Experiment 3a and 3b, we found that when categories were considered singly, participants did, in fact, generate more convincing explanations for high as compared with low coherence categories. And, in Experiment 4, when multiple categories were available, participants were more likely to include the high coherence category in their explanations. These results are interesting in their own right in that they further support the more general claim that high and low coherence categories are treated differently in reasoning.

Reasoning from multiple categories

The present research concludes that category coherence influences reasoning about cross-classified entities. However, one potential concern is that, because coherence was found in the pretest to be correlated with similarity, it was actually similarity driving the results. Such a correlation often arises because similarity judgments are based on both common features among category members and common relations among the features (Lassaline, 1996; Markman & Gentner, 1993; Rehder & Burnett, 2005; Rehder & Hastie, 2001). It is reasonable to state that similarity influences reasoning results, as long as it is understood that judgments of similarity are heavily influenced by common relations among features, rather than just common features. What is not clear from the present work is the extent to which categories that have only relations in common (e.g., ad hoc categories; Barsalou, 1983) would also be deemed more useful for inference than those with only features in common.

We found no evidence that distinctiveness in the population in and of itself plays a role in category-based inference in the manner predicted by Nelson and Miller (1995). On the other hand, high distinctiveness might emerge in a number of ways, at least one of which is associated with coherence. In particular, a category might be distinctive as a result of having one or more deep properties that few other entities possess (e.g., sky divers being risk seeking). In contrast, a category might be distinctive merely as a result of an environment's low demand for members (e.g., county clerks). Because of its relationship with coherence, the first kind of distinctiveness would be expected to promote inference while the second kind would not.

However, Nelson and Klutas (2000) operationalized distinctiveness in yet a different way, as the relative number of category members in a particular context rather than in the population at large (e.g., a county clerk in a roomful of lawyers), and their finding that participants reasoned from the more distinctive category cannot easily be explained by our coherence account.

Our Experiment 2 touched on a different issue in cross-classification—the extent to which a probability judgment is influenced by more than one category. Murphy and Ross (1999, Experiments 4 and 5) contrasted food categories that were of different types (e.g., taxonomic versus script-based categories), rather than social categories of the same type, using the methodology that we adopted for Experiment 2. They found that, when two categories were consulted, only 30% of the time was a judgment given anywhere in between the two category probabilities; the remainder of the time the judgment exactly matched one of the two category probabilities. If we adopt their scoring and define an intermediate judgment to include all values in between the two probabilities, judgments in our study fell between the two categories 82% of the time.

That participants were more inclined to integrate information here than in Murphy and Ross (1999) might be due to the fact that all categories here were of the same type and, thus, were equally relevant to the property in question. This is in contrast with past work, where categories were of different types within a problem. While our result is not intuitively surprising, it is important given that past work in this area has identified many contexts in which, once a "best" category has been selected, other available categories are not consulted in the formation of a judgment (Murphy & Ross, 1999;

Ross & Murphy, 1999). The present work identifies one situation in which multiple categories are used. Overall, however, the point remains that reliance on multiple categories seems to be underutilized as a strategy and reserved for very particular situations in which multiple equally relevant categories are available.

Coherence and category-based explanations

Work by Lassaline (1996) and Rehder and colleagues (Rehder & Burnett, 2005; Rehder & Hastie, 2001) on category-based inference has focused on local coherence around a particular property, not on overall category coherence. Rehder and colleagues (Rehder & Burnett, 2005; Rehder & Hastie, 2001) found that people are more likely to transfer a property from a category to a new exemplar to the extent that the causal antecedent of the property in question is present in the new exemplar. For example, if category members are known to waddle because they have large bodies, then an inference about whether or not a new category member waddles would be strongest if the individual were known to also have a large body.

In the present work, because we used novel properties with no explicit links to known category features, we cannot speak to questions about the local coherence around an individual property. However, what we do know is that even when no causal link is provided from known category features to a novel property, the higher coherence category is still preferred as the basis for inference. One possibility is that higher coherence categories offer a greater number of deep features on which to base plausible causal explanations, so they might more often be used as the source of explanation and thus as the source of inference as well. This possibility is suggested by the results of our last two experiments, and is consistent with Sloman (1994) who found an increased likelihood to transfer a property from one category to another if the same explanation could be used to account for the presence of the property in each case.

Of course, one does not necessarily have to actually generate causal explanations in order to do a basic inference task. One could instead just estimate the likelihood of there being a good explanation from category coherence. A benefit of the explanation account, however, is that it also provides for situations in which an inference is made from a low coherence category. This usually occurs when the property in question is more related to the low coherence category. For example, if skydivers tend to have poor short term memory while waiters tend to have good short term memory, and an individual is a member of both categories, one might infer that the individual has a good memory because this is essential to remembering restaurant orders but is not essential (to our knowledge) to skydiving. Further work is needed to explore these possibilities.

Nature of coherence in social categories

Assessing the coherence levels of natural social categories raises many questions about the ways in which categories are coherent and how different kinds of coherence might influence reasoning. The results of our pretesting showed that the job and hobby categories used here were all at least somewhat coherent, as one would expect of categories in everyday use. Nonetheless, some of these categories were still more coherent than others. Given that we used established coherence measures, the results suggest that the way coherence is manifested in job and hobby social categories is not unlike the way it is manifested in other social categories (Haslam et al., 2000), natural kinds, and artifacts (Lassaline, 1996; Rehder & Burnett, 2005; Rehder & Hastie, 2001). While these measures do not capture more subtle differences in the ways in which categories might be coherent, they provide a useful start to bridging the gap between existing data and natural social categories.

What is unique to the categories studied here is the nature of category features. Past work suggests that, unlike other categories, social categories appear to be dominated by personality traits. In particular, Dahlgren (1985) found that, when asked for the definitions of job categories (e.g., doctors), participants generated as many personality trait features (“is intelligent”) as behavioral (“attends to people healing”) or relational (“works with other doctors”) ones, and few perceptual features (“wears a stethoscope”). In light of Dahlgren’s (1985) findings, we revisited our own data (the data were recoded by one of the authors) and found that ≈50% personality traits, 45% behaviors and relations, and 5% perceptual features were generated in response to the pretested categories. The findings are surprising given that job and hobby categories could have easily been described exclusively in terms of activity-related behaviors.

We speculate that activity-based categories such as jobs and hobbies are represented at one level as a set of behaviors but, at an underlying level, as a set of personality traits that give rise to the behaviors (see Wattenmaker, 1995; Yulill, 1992, for similar suggestions). For example, at one level, skydivers are people who routinely jump out of planes with a parachute. At another level, skydivers are typically people with a high need for excitement who seek out risky activities such as skydiving. While this characterization might not accurately describe all skydivers, it is a very useful feature level for making far-reaching inferences about skydivers (even beyond the domain of skydiving). It is also consistent with the well-known fundamental attribution error, the phenomenon whereby people have a tendency to attribute behaviors to personality traits more often than to situational constraints (Ross, 1977).

It follows that high versus low coherence categories might differ in the extent to which personality traits are central to the category representation, with high coherence categories being associated with more underlying traits. That there might be differences across categories in the development of underlying traits would not be surprising. It readily comes to mind for most people, for example, that the desire to be a skydiver comes from a need for excitement. A trait-based explanation for becoming a county clerk is more difficult to generate and might even lead to the conclusion that county clerks are such for situational rather than personality-oriented reasons. This speculation requires future empirical substantiation, such as through experiments using novel social categories or through experiments exploring the goodness of category members missing trait properties (e.g., a risk-averse skydiver).

Conclusions

Category-based induction involves not only assigning an entity to one or more categories but also deciding which of these categories to use to inform inference. Based on the experiments presented in this paper, we conclude that natural social categories vary in coherence, the coherence of social categories is an important determinant of which one or more categories are selected and used to make an inference, and category-based explanation may serve as an important mechanism for linking novel properties to existing relational structures in the service of assessing the strength of an inference. This work complements a growing body of research on category coherence and category-based inference.

Appendix A. Materials used in Experiment 1

Similarity ratings are on a scale from 1 (Not at all similar) to 7 (Highly similar). Frequency estimates are per 1000 people. To reduce the influence of extreme values, individual-participant frequency estimates were log-transformed before each category mean was computed and then reverse-transformed.

Item A	Sim	Freq	Item B	Sim	Freq	Property (reversed for Item B)
Items matched on frequency estimates						
Soldier	4.3	32	Matchbook collector	2.9	17	Terriers to beagles
Feminist supporter	4.5	55	Waiter	2.3	73	The color red to blue
Minister	4.9	12	Rubber-stamp collector	3.1	11	Pepsi to Coca-Cola
Pro wrestler	5.4	3	Badminton player	2.4	14	Reading fiction to non-fiction
Yacht club member	4.7	16	County clerk	2.8	10	Chinese to Mexican food
Rare-sculpture collector	4.6	6	Limousine driver	3.1	13	Comedy movies to adventures
M	4.7	21	*M*	2.8	23	
Items matched on similarity ratings						
Charity volunteer	3.3	86	Amateur bassoon player	3.5	6	Basketball to football
Labor union member	3.1	213	Carpet manufacturer	3.3	13	Drinking gin to whiskey
Feminist supporter	4.5	55	Opera singer	4.5	3	Daffodils to tulips
Police officer	4.1	40	Rare-sculpture collector	4.6	6	Vacationing in Bermuda to Bahamas
Sports car owner	3.1	66	Rubber-stamp collector	3.1	11	Mandarin to Cantonese
Science fiction addict	4.1	56	French chef	4.1	4	Watching ABC to NBC
M	3.7	86	*M*	3.9	7	

Appendix B. Entitativity scales

Labels (in bold) were not seen by participants. Inherence responses were reverse coded.

B.1. Uniformity

"Some categories contain members who are very similar to one another; they have many things in common. Members of these categories are relatively uniform. Other categories contain members who differ greatly from one another, and do not share many characteristics." Endpoints: 1 (Diverse/Differing) to 9 (Uniform/Similar).

B.2. Informativeness

"Some categories allow people to make many judgments about their members; knowing that someone belongs in the category tells us a lot about that person. Other categories only allow a few judgments about their members; knowledge of membership is not very informative." Endpoints: 1 (Uninformative/Few judgments) to 9 (Informative/Many judgments).

B.3. Inherence

"Some categories have an underlying reality; although their members have similarities and differences on the surface, underneath they are basically the same. Other categories also have similarities and differences on the surface, but they do not correspond to an underlying reality." Endpoints: 1(Underlying reality or sameness) to 9 (No underlying reality or sameness).

Appendix C. Materials used in Experiment 2

Letters "A" and "B" in headings refer to Versions A and B of materials. "P" refers to problem number. Categories within each problem are listed below from most to least coherent (H = high, MH = medium–high; ML = medium–low; L = low). Similarity ratings are on a scale from 1 (Not at all similar) to 7 (Highly similar). "Pos" refers to the position of the category during presentation. And "Perc" refers to the percentage of category members that participants were told had the given property.

P	Category	Sim	Pos A/B	Perc A/B (%)	Property A/B
1	Minister (H)	4.9	4/1	55/30	Vacationing in Bermuda to Bahamas/reading non-fiction to fiction
	Rare-sculpture collector (MH)	4.6	3/3	40/20	
	Chimney cleaner (ML)	3.5	2/2	30/55	
	Weekend badminton player (L)	2.4	1/4	20/40	
2	Country club member (H)	5.0	2/3	90/45	Drinking gin to whiskey/beagles to terriers
	Trial lawyer (MH)	4.2	4/1	80/55	
	Sports car owner (ML)	3.1	1/2	45/90	
	Amateur tennis player (L)	2.6	3/4	55/80	
3	Professional wrestler (H)	5.4	3/1	75/15	Reading fiction to non-fiction/drinking Pepsi to Coca-Cola
	Sky diver (MH)	3.6	1/2	85/20	
	Museum guard (ML)	3.3	4/3	15/75	
	Matchbook collector (L)	2.9	2/4	20/85	
4	Yacht club member (H)	4.7	1/2	20/55	The sound of Mozart to Beethoven/the color blue to red
	French chef (MH)	4.1	2/4	25/40	
	Spice gardener (ML)	3.6	3/1	55/20	
	Notary public (L)	2.9	4/3	40/25	
5	Feminist supporter (H)	4.5	2/4	30/85	Terriers to beagles /the sound of Mozart to Beethoven
	Science fiction addict (MH)	4.1	3/2	10/80	
	Charity volunteer (ML)	3.3	4/3	85/30	
	Waitress (L)	2.3	1/1	80/10	
6	Professional opera singer (H)	4.5	4/4	60/70	Mexican food to Chinese food/comedies to adventure movies
	Amateur bassoon player (MH)	3.5	2/3	55/90	
	Rubber-stamp collector (ML)	3.1	1/2	70/60	
	Lottery winner (L)	2.0	3/1	90/55	
7	Medieval studies major (H)	4.8	3/2	40/10	The color red to blue/Chinese food to Mexican food
	Houseboat owner (MH)	4.0	1/3	60/25	
	Amateur actress (ML)	2.8	2/4	10/40	
	Hotel clerk (L)	2.7	4/1	25/60	
8	Police officer (H)	4.1	1/4	15/85	Football to basketball/drinking whiskey to gin
	Hunter (MH)	3.7	4/2	20/75	
	Snake owner (ML)	3.0	3/1	85/15	
	Cashier (L)	2.1	2/3	75/20	
9	Wine-of-month club member (H)	5.0	2/3	70/30	Adventure movies to comedies/vacationing in Bahamas to Bermuda
	Amateur tango dancer (MH)	4.0	4/1	85/10	
	Water polo fan (ML)	3.6	1/4	30/70	
	County clerk (L)	2.8	3/2	10/85	
10	Protestant (H)	3.6	3/1	55/75	Drinking Coca-Cola to Pepsi/basketball to football

Appendix C (*continued*)

P	Category	Sim	Pos A/B	Perc A/B (%)	Property A/B
	Labor union member (MH)	3.1	1/4	45/90	
	Dog owner (ML)	2.3	4/3	75/55	
	Registered voter (L)	1.6	2/2	90/45	
Means	High coherence (H)	4.7			
	Medium–high coherence (MH)	3.9			
	Medium–low coherence (ML)	3.2			
	Low coherence (L)	2.4			

Appendix D. Materials used in Experiment 4

Similarity ratings are on a scale from 1 (Not at all similar) to 7 (Highly similar). Frequency estimates are per 1000 people. To reduce the influence of extreme values, individual-participant frequency estimates were log-transformed before each category mean was computed and then reverse-transformed.

Item A (High coh)	Sim	Freq	Item B (Low coh)	Sim	Freq	Property (reversed for item B)
Soldier	4.3	32	Matchbook collector	2.9	17	Terriers to beagles
Feminist supporter	4.5	55	Waiter	2.3	73	The color red to blue
Minister	4.9	12	Rubber-stamp collector	3.1	11	Pepsi to Coca-Cola
Pro wrestler	5.4	3	Badminton player	2.4	14	Reading fiction to non-fiction
Yacht club member	4.7	16	County clerk	2.8	10	Chinese to Mexican food
Rare-sculpture collector	4.6	6	Limousine driver	3.1	13	Comedy movies to adventures
Hunter	3.7	98	Cashier	2.1	103	Mozart to Beethoven
Trial lawyer	4.2	24	Amateur tennis player	2.6	35	VISA to MasterCard
Brain surgeon	6.6	3	Amateur quilter	3.9	2	Gin to whiskey
Police officer	4.1	40	Water polo fan	3.6	52	Honda to Toyota
Weekend sky diver	3.6	20	Hotel clerk	2.7	17	Newsweek to Time magazine
Science fiction addict	4.1	56	Museum guard	3.3	8	Watching NBC to ABC
M	4.6	30	*M*	2.9	30	

References

Ahn, W. (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., & Kim, N. S. (2000). The causal status effect in categorization: An overview. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (pp. 23–65). San Diego: Academic Press.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.

Atran, S. (1990). *Cognitive foundations of natural history*. New York: Cambridge University Press.

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

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as a determinant of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629–654.

Dahlgren, K. (1985). The cognitive structure of social categories. *Cognitive Science, 9*, 379–398.

Erickson, J. E., Chin-Parker, S., & Ross, B. H. (2005). Inference and classification learning of abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 86–99.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*, 113–127.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*, 569–592.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (pp. 163–199). San Diego: Academic Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 411–422.

Hirschfeld, L. A. (1994). Is the acquisition of social categories based on domain-specific competence or knowledge transfer? In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind* (pp. 201–233). Cambridge, England: Cambridge University Press.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and child's construction of human kinds*. Cambridge, MA: MIT Press.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development, 63*, 1536–1557.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 829–846.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 754–770.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3–28.

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

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23*, 250–269.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 421–467.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469–1481.

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.

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

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 904–919.

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

Murphy, G. L., & Ross, B. H. (1999). Induction with cross-classified categories. *Memory & Cognition, 27*, 1024–1041.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberchien (Ed.), *Advances in cognitive science, Vol. 2: Theory and applications* (pp. 23–45). Chichester, England: Ellis Horwood.

Nelson, L. J., & Klutas, K. (2000). The distinctiveness effect in social interaction: Creation of a self-fulfilling prophecy. *Personality and Social Psychology Bulletin, 26*, 126–135.

Nelson, L. J., & Miller, D. T. (1995). The distinctiveness effect in social categorization: You are what makes you unusual. *Psychological Science, 6*, 246–249.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 416–432.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264–314.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, 323–360.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91*, 113–153.

Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261–1275.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495–553.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.). *Advances in Experimental Social Psychology* (Vol. 10). New York: Academic Press.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction, and social cognition* (pp. 11–36). London: Sage.

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885–904.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence judgments on likelihood. *Cognition, 52*, 1–21.

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

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology, 71*, 888–898.

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

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

Spellman, B. A., Lopez, A., & Smith, E. E. (1999). Hypothesis testing: Strategy selection for generalizing versus limiting hypotheses. *Thinking and Reasoning, 5*, 67–91.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

van Rijswijk, W., & Ellemers, N. (2002). Context effects on the application of stereotype content to multiple categorizable targets. *Personality and Social Psychology Bulletin, 28*, 90–101.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology, 28*, 274–328.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18*, 158–194.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 449–468.

Yulill, N. (1992). Children's conception of personality traits. *Human Development, 35*, 265–279.