Patalano, A. L., Ross, B. H., & Chin-Parker, S. (2003). The role of coherence in category-based explanation. In *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* [refereed] (pp. 910-915). Mahwah, NJ: Erlbaum.

The Role of Coherence in Category-Based Explanation

Andrea L. Patalano (apatalano@wesleyan.edu)
Department of Psychology, Wesleyan University, 207 High Street
Middletown, CT 06459 USA

Seth Chin-Parker (chinpark@s.psych.uiuc.edu)
Department of Psychology, University of Illinois, 603 East Daniel Street
Champaign, IL 61820 USA

Brian H. Ross (bross@s.psych.uiuc.edu)
Beckman Institute, University of Illinois, 405 North Mathews Avenue
Urbana, IL 61801 USA

Abstract

Category coherence refers to the extent to which a category is perceived to be a meaningful whole (Patalano & Ross, 2002; Ross & Patalano, 2002). We tested the hypothesis that category coherence influences the extent to which a category is used in the generation of category-based causal explanations of social behavior and preferences. In Experiments 1a and b, participants were told that members of a category shared a particular preference (e.g., sky divers prefer fiction to non-fiction), and were asked to generate the most plausible explanation for the preference. Explanations generated for high coherence categories were more plausible than those generated for low coherence categories. In Experiment 2, high and low coherence categories were contrasted in the context of a single problem. Participants were told that members of two categories (e.g., people who are both sky divers and pianists) shared a particular preference and were again asked for the most plausible explanation of the preference. References to the high coherence category occurred more often than those to the low coherence category. It is concluded that coherence influences both category selection and quality of category-based causal explanation. Implications of this work and future research directions are discussed.

Introduction

We are constantly engaged in trying to make sense of the world around us. Towards this goal, we rely on multiple kinds of causal explanations for events, behaviors, properties, etc. For example, sometimes we attribute human behavior to a situation (e.g., "the picketers' actions were caused by unfair management"). At other times we attribute it to a personality trait (e.g., "the donors gave money to the church because they were kind-hearted"). And at still other times we attribute it to a category membership (e.g., "the concert goers enjoyed the loud music because they were teenagers"). Developing and testing our theories allows us to integrate knowledge, to generate predictions and expectations, and to acquire a deeper understanding of how our environment works.

Category-based explanations may play a particularly important role in social reasoning. People are frequently described in terms of social categories, making these categories a salient source of potentially useful information. One challenge to explanation in this context, however, is that most people belong to multiple social categories. So it is difficult to know to which of the multiple categories to attribute a behavior or preference in question. For example, imagine noticing that a group of people who happen to be both war veterans and Italian-American immigrants are very patriotic. Is this because war veterans are patriotic as a result of military experience? Or is it because certain groups of immigrants may feel patriotic as a result of their immigrant experience? Because entities and events often belong to multiple categories, developing category-based explanations must rely on some cognitive heuristics or strategies for relevant explanatory categories.

One obvious way that people might choose relevant explanatory categories is by looking for pre-existing relationships between the property in question and known categories. For example, if I already know for certain that veterans are patriotic because the military works to instill patriotism in its soldiers, then I will likely rely on this knowledge. In this case, I may simply be retrieving the explanation. But what about situations in which people are striving to explain a newly discovered property or to construct a novel explanation for a pattern of behavior? Are there any properties of categories that might lead these categories to be used over others?

One answer to this question is to think about what is needed of a category in order for it to be useful for category-based explanation. First, use of a category requires considerable knowledge about the category so that novel connections might be developed between the category and the to-be-explained behavior. At least, the more knowledge is available, the more likely a plausible connection might be found. Second, explanation relies not just on knowledge of surface features of categories, but also on knowledge of deeper causal properties that could give rise to a host of surface properties (e.g., Chi, Feltovich, & Glaser, 1981). For example, the category "sky diver" suggests the property of being "risk-seeking" which may explain many other behaviors and preferences of members of this category.

Third, explanation works best when a category has little within-category variability, so that one can be reasonably confident that a property associated with the category applies to many of its members.

Some of these identified characteristics have, in fact, already been studied in the context of understanding structural differences in social categories. Haslam and colleagues (Haslam, Rothschild, & Ernst, 2000), for example, asked participants to rate social categories on twelve properties, including properties similar to those identified in the previous paragraph. A factor analysis revealed a cluster of associated properties, which the authors referred to as an *entitativity* factor, which included: *uniformity* (degree of within-category similarity of category members), *informativeness* (extent to which the category provides information about its members), and *inherence* (the extent to which the category is associated with deep, underlying features).

Because the term "entitativity" has multiple meanings in the social psychology literature, we will use the alternative term *coherence* here to refer to the extent to which a category is treated as a meaningful whole. The notion of coherence is connected with a theory based view of concepts (Murphy & Medin, 1985) in that coherence depends on an understanding of the category that goes beyond an enumeration of category properties (as in a prototype structure; e.g, Medin & Smith, 1984), or an unanalyzed storage of exemplars in memory (e.g., Medin & Schaffer, 1978).

There is little research on the relationship between category coherence and the use of categories in explanation. This is an important topic for cognitive science because explanation is known to play a central role in education and self-discovery (e.g., Chi, de Leeuw, Chiu, & LaVancher, 1994), machine learning (e.g., DeJong & Mooney, 1976), and knowledge representation and use (e.g., Keil & Wilson, 1995). The topic is also central to philosophical thought on the nature of causality and explanation (e.g., Pettit, 1995), as well as to cross-cultural approaches to reasoning (e.g., Lopez, Atran, Coley, & Medin, 1997).

The goal in conducting the following two experiments was to investigate the extent to which category coherence predicts the quality of category-based causal explanations (Experiment 1), and therefore the relative use of a category for causal explanation when multiple categories are available (Experiment 2).

Experiments 1a and b

Experiment 1a

In this experiment, 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 high and low in coherence were selected from a database of social categories that had been previously rated by University of Illinois undergraduates (Patalano & Ross, 2002; Ross & Patalano, 2002). These categories were rated, using Likert scales, on the previously-described coherence dimensions of uniformity, informativeness, and inherence, as well as a related scale called *similarity* (which taps into the same property as uniformity). Because correlations between pairs of dimensions were high (r>.93), the dimension of similarity was selected as an estimate of coherence for creating the materials used in Experiments 1 and 2. However, the same categories could have been selected using any or all of the other dimensions.

After completing the problems, participants rated each explanation for plausibility. We hypothesized that high coherence category explanations would be rated as more plausible than low coherence ones.

Method

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

Materials Each booklet consisted of 12 problems, 6 with high coherence categories and 6 with low coherence categories. The problems were of the following format (professional wrestlers, the category used in this example, is a high coherence category):

> 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.

Categories high versus low in coherence were selected from a database of categories for which coherence ratings had been previously obtained (Patalano & Ross, 2002; Ross & Patalano, 2002). For the experiments in this paper, coherence was estimated from a single coherence measure called *similarity*. Similarity refers to "the similarity of two randomly selected category members to one another" where a rating of 1 corresponds to "Not at all similar" and a 7 corresponds to "Highly similar." In Experiment 1, the similarity ratings for the high coherence categories ranged from 4.3-5.4 with a mean of 4.7; the ratings for the low coherence categories ranged from 2.3-3.1 with a mean of 2.8 (see Appendix for materials).

The following two additional constraints were placed on category selection. First, categories were selected so that half of the categories at each coherence level were occupation categories and half were hobby categories. Second, the estimated frequencies of category members per 1000 people (also in the previously-mentioned

database) were matched across high and low coherence categories. On average, the estimated frequency of high coherence category members was 21/1000 versus 23/1000 for low coherence ones. These two constraints assured that coherence was not confounded with category type or estimated category frequency in this experiment.

The properties used in each problem (e.g., preferring Bermuda versus the Bahamas), also shown in the Appendix, were chosen to be relatively "blank" in the sense that would not have been previously associated with the problem category.

Two versions of the booklets were made, with different random orders of problems and different pairings of categories and properties in each version.

Procedure Participants were tested in a group in a 30 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).

Results

The results from the two booklet versions showed the same pattern and so the data were collapsed. Every participant generated an explanation and a rating for each problem. For example, a response given by one participant to the high-coherence wrestler example used earlier was "Wrestlers are more daring and want to go to dangerous, risky areas such as Bermuda [over the Bahamas]" (given a plausibility rating of 4). A response by the same participant to the low-coherence rubber-stamp collector category (whose 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]" (given a plausibility rating of 2).

The mean plausibility rating for high coherence categories was 3.83 (SD=0.93) versus 3.10 (SD=1.00) for low coherence ones, $t(7)=2.44$, $p=.04$. The difference in magnitude is 0.73 Likert-scale points. In other words, participants generated better (in their opinions) explanations for high as compared with low coherence categories.

Experiment 1b

It is possible that plausibility ratings in Experiment 1a were influenced by participants having actually generated the explanations themselves. While it is not clear how this could have lead to different ratings for the two kinds of categories, Experiment 1b addresses this potential problem. In this experiment, each of the eight completed booklets of Experiment 1a (except for the plausibility ratings) was given to new participants. These participants were asked to assign a plausibility rating for each explanation in their booklet. We expected the plausibility ratings generated by these participants to show the same pattern of results as those of Experiment 1a.

Method

Participants Twenty four undergraduates at the University of Illinois participated in exchange for introductory course credit.

Materials The explanations from the 8 booklets from Experiment 1a were typed into 8 new booklets (simply so that the new participants would see typed rather than hand-written explanations), and the earlier plausibility ratings were omitted.

Procedure Each of the 8 booklets was given to 3 participants. The participants were asked to go through the booklets, to read each explanation, and to rate each on a scale from 1 (Highly implausible) to 7 (Highly plausible). Participants were tested in groups of 8 in 30 min sessions.

Results

As in Experiment 1a, the same pattern of results was found for both versions of the booklets so the data were collapsed. The high coherence category explanations (M=3.95, SD=.74) were again rated as being more plausible than the low coherence category ones (M=3.36, SD=.79), $t(23)=4.01$, $p<.001$. The difference in magnitude is 0.59 Likert-scale points.

Discussion

The goal of Experiment 1 was to investigate the extent to which category coherence predicts quality of category-based causal explanations of preferences. The results provide evidence in support of the hypothesis that people generate better causal explanations for high coherence as compared with low coherence categories.

The order of magnitude of this effect is, on average, 0.66 Likert-scale points. While this effect may appear somewhat small in size, it should be considered in the context of the following three qualifications. 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 is related to the goal of Experiment 2. Recall that the motivation for these studies, as described in the introduction, is to understand the role of category coherence in category-based explanation for preferences. It was hypothesized that, when multiple categories are

available, people may be more inclined to reason from higher coherence categories. The first experiment provided evidence that coherence is in fact related to perceived plausibility of category-based explanations. The second experiment considered whether or not relative coherence predicts use of one category over another in explaining preferences.

Experiment 2

In this experiment, each problem made reference to people who were members of two categories (one high coherence and one low coherence) and had a novel hypothetical preference. 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.

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.

Categories high versus low in coherence were selected from a database of categories as in Experiment 1. The 12 categories used in Experiment 1 were paired here to create 6 problems; 6 more problem were created using new categories. The similarity ratings for the high coherence categories ranged from 3.6-6.6 with a mean of 4.6; the ratings for the low coherence categories ranged from 2.1-3.9 with a mean of 2.6. Though the distributions overlapped (it was not possible to create non-overlapping sets, using the existing database, with the additional constraints given below), categories in the same problem differed by at least 0.5 points in the right direction.

As in Experiment 1, high and low coherence category sets were matched on number of job versus hobby categories (half of each) and on estimated frequency of category members per 1000 people. Overall, the average frequency was 30 people per 1000 for both high and low coherence categories. In addition to equating frequency across the categories, individual-problem category pairs were approximately matched on frequency as well.

The properties used in each problem (e.g., preferring Bermuda versus the Bahamas) were again chosen to be relatively "blank" in the sense that they would not have been previously associated with the problem category.

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: *Hi-Coh* (explanation makes reference only to the high coherence category, *Lo-Coh* (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 told 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. It should nonetheless be noted that the results would not change in pattern if either one or the other of the coder's responses, rather than both of them, had been used.

Summary values for each participant were computed both (1) for all explanations (*All-Explanations* analysis), and (2) for the most plausible explanations only (one per problem, as identified by participants; *Plausible-Only* analysis). Summary values were computed by simply adding the points given to each code for that participant. The total number of points available to be divided among the codes was 36 (12 problems x 3 explanations per problem) for the All-Explanations analysis, and 12 (12 problems x 1 explanation per problem) for the Plausible-Only analysis.

For intuitiveness of understanding, the points for each code for each participant were then recomputed as a

percentage of the total number of points. Percentages technically refer to relative points assigned to a category rather than relative use of the category. These are slightly different things given that coders occasionally disagreed on how to code an explanation resulting in points being divided over two codes. However, because discrepancies occurred on such a small number of occasions (21 out of 648), for ease of discussion, no distinction will be made between these two interpretations of the percentages.

Analyses The results of the All-Explanations analysis will be described first. As illustrated in Table 1, for nearly half of the explanations (49%), participants made explicit reference to neither category. In these cases, participants generally ignored the base rate information (e.g., that vacations in Bermuda and the Bahamas are equally popular in the population at large) 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 drawing causal explanations from nearly blank 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). This finding is consistent with other evidence suggesting that people often do not integrate multiple categories in reasoning (e.g., Malt, Ross, & Murphy, 1995).

Of central importance to the present investigation is 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 (for 22% of all explanations; 56% for this subset of the data) reliably more often than the low coherence categories (for 17% of all explanations; 44% for this subset of the data; $t(17)=2.58$, $p=.02$).

The results for the Plausible-Only analysis follow the same pattern but show even greater reliance on high as compared with low coherence categories. Again, as illustrated in Table 1, 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. We found that participants relied on the high coherence categories (for 22% of all most-plausible explanations; 67% for this subset of the data) twice as often as the low coherence categories (for 11% of all most-plausible explanations; 33% for this subset of the data; $t(17)=2.62$, $p=.02$). 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 1: Categories Used in Explanations

	Hi-Coh	Lo-Coh	Both	Neither
All Explanations	22%(13)	17%(17)	12%(13)	49%(32)
Plausible Only	22%(14)	11%(11)	14%(20)	53%(32)

Note: Standard deviations are in parentheses.

Discussion

The goal of this experiment was to assess the extent to which category coherence influences category use in causal 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 causal 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 Experiment 1 in which it was found that more-plausible explanations were generated for high coherence as compared with low coherence categories. The studies taken together suggest that high coherence categories are more often used in causal explanation precisely because these categories afford generation of plausible explanations.

General Discussion

The experiments described here provide an initial understanding of one aspect of category structure – category coherence – that influences category selection in the service of category-based causal explanation. Coherence is important in that high coherence categories are used more often than low coherence categories for generating explanations, and in that the explanations generated for high coherence categories are more plausible ones.

This work also provides evidence for the validity of the construct of category coherence. Past work has shown that people consistently rate some categories as high in coherence and others as low (as measured by uniformity, similarity, informativeness, and inherence scales). However, there has been little work showing the actual influence of coherence on reasoning. In addition to the work described here, we are exploring the role of coherence in other category-based reasoning tasks such as induction and generalization.

In related research, we are also beginning to do a more refined analysis of the content and structure of mental representations of high versus low coherence categories. This work will allow us to provide further evidence for the greater inter-connectedness of deep and surface properties in high coherence categories. A careful analysis of the content of various categories could allow us to better understand

coherence and to better understand how causal explanations use category knowledge for high versus low coherence categories.

This current work raises many other questions as well regarding the mechanism by which coherence influences category selection in explanation, how category information interacts with other kinds of information (e.g., personality, situational, etc.) in the service of aiding explanation, and how coherence influence what is learned about categories of various coherence levels as a result of category use in explanation.

Acknowledgments

This work was funded in part by NSF grant SBR 97-20304 to Brian H. Ross. We thank Jane Erickson for her assistance in conducting these studies.

References

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-125.

Chi, M. H. T., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science, 18(3),* 439-447.

DeJong, G., & Mooney, R. J. (1986). Explanation-based learning: An alternative view. *Machine Learning, 1(2),* 145-176.

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

Keil, F. C., & Wilson, R. A. (Eds). (2000). *Explanation and Cognition.* Cambridge, MA: MIT Press.

Lopez, A., Atran, S., Coley, J. D., & Medin, D. L. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32(3),* 251-295.

Malt, B. C, Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21(3),* 646-661.

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

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Anuual Review of Psychology, 35,* 113-138.

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

Patalano, A. L., & Ross, B. H. (2002, June). *The role of coherence in reasoning about individuals belonging to multiple categories.* Poster presented at the Fifth Annual Meeting of the American Psychological Society, New Orleans, Louisiana.

Pettit, P. (1995). Causality at higher levels. In D. Sperber, D. Premack, & R. Premack (Eds.), *Causal Cognition* (pp. 399-422). New York: Oxford University Press.

Ross, B. H., & Patalano, A. L., (2002, November). *Category-based inference from cross-categorized items.* Paper presented at the Forty-Third Annual Meeting of the Psychonomic Society, Kansas City, Missouri.

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

Appendix: Materials Used in Experiment 1

Coh refers to coherence as estimated by pretest participants' mean ratings of with-category similarity on a scale from 1 (low similarity) to 7 (high similarity). Frequency refers to pretest participants' estimated frequency per 1000 people in the United Stated. Property sets used in Version A of the materials are listed in parentheses. Version B used a different pairing of the same categories and properties. One category and property set was used for each problem in Experiment 1.

Hi Coherence Categories	**Coh**	**Freq**	**Lo Coherence Categories**	**Coh**	**Freq**
1. soldier (gin/whiskey)	4.3	32	7. matchbook collector (terrier/beagle)	2.9	17
2. feminist supporter (red/blue)	4.5	55	8. waiter (football/basketball)	2.3	73
3. minister (Coke/Pepsi)	4.9	12	9. rubber-stamp collector (tulips/roses)	3.1	11
4. pro wrestler (fiction/non-fiction)	5.4	3	10. badminton player (fiction/non)	2.4	14
5. yacht club member (Mex/Chin food)	4.7	16	11. county clerk (Mandarin/Cantonese)	2.8	10
6. rare-sculpture collector (NBC/ABC)	4.6	6	12. limo driver (adventures/comedies)	3.1	13
M=	*4.7*	*21*	*M=*	*2.8*	*23*