Cognition 110 (2009) 248–253

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Explanation and categorization: How "why?" informs "what?"

Tania Lombrozo *

Department of Psychology, UC Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 9 June 2008
Revised 7 October 2008
Accepted 10 October 2008

Keywords:
Categorization
Feature centrality
Explanation
Causation
Causal status
Functions
Functional explanation
Teleological explanation

ABSTRACT

Recent theoretical and empirical work suggests that explanation and categorization are intimately related. This paper explores the hypothesis that explanations can help structure conceptual representations, and thereby influence the relative importance of features in categorization decisions. In particular, features may be differentially important depending on the role they play in explaining other features or aspects of category membership. Two experiments manipulate whether a feature is explained mechanistically, by appeal to proximate causes, or functionally, by appeal to a function or goal. Explanation type has a significant impact on the relative importance of features in subsequent categorization judgments, with functional explanations reversing previously documented effects of 'causal status'. The findings suggest that a feature's explanatory importance can impact categorization, and that explanatory relationships, in addition to causal relationships, are critical to understanding conceptual representation.

Published by Elsevier B.V.

1. Introduction

Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.
–Lewis Carroll (1871)

Why are toves slithy? One way to explain the slithiness of toves is *mechanistically*, in terms of a proximate causal mechanism. Just as tigers' stripes can be explained by appeal to underlying pigments, perhaps the slithiness of toves can be explained by appeal to a substance in tove's diet. Another way to explain the slithiness of toves is *functionally*, in terms of a function or goal. Just as tigers' stripes can be explained by appeal to camouflage, perhaps toves' slithiness serves an important purpose, such as gimbling in the wabe.

Lewis Carroll's whimsical creature illustrates the generality of mechanistic and functional explanations. Any feature with a function, such as a biological adaptation or the component of an artifact, will typically support a mechanistic explanation in terms of proximate causes as well as a functional explanation in terms of a function or goal. This paper explores the hypothesis that different kinds of explanations reflect deep differences in reasoning with consequence for categorization. In particular, mechanistic explanations may reflect reasoning in terms of physical mechanisms, akin to Daniel Dennett's "physical stance," while functional explanations may reflect reasoning in terms of functions and goals, akin to Dennett's "design stance" (Dennett, 1987; see also Keil, 1994).

A pluralistic approach to explanation and reasoning is attractive because causal systems support different kinds of generalizations. Artifacts and biological adaptations typically support some generalizations best captured in terms of physical mechanisms, and others best captured by functions and design. To explain or predict what happens when a computer's power button is depressed, a design stance will do well. To explain or predict what happens if a computer is run in a magnetic field, a physical stance will do better. Which stance is most appropriate depends on the system and judgment in question. There is no "all-purpose" stance, just as there is no "all-purpose" explanation that addresses every aspect of a "why?" question.

* Tel.: +1 510 642 7145; fax: +1 510 642 5293.
E-mail address: lombrozo@berkeley.edu.

0010-0277/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.cognition.2008.10.007

Taking pluralism further, one might expect category membership to likewise depend on the system and judgment in question. Category representations are typically posited in the service of inferential utility: they provide "maximum information with the least cognitive effort" (Rosch, 1999, p. 190), capitalizing on "information-rich cluster[s] of attributes in the environment" (Rosch, 1999, p. 197). Just as different stances can prove more or less useful in understanding a given system for a given purpose, different aspects of these "information-rich clusters" may be more or less useful depending on the reason for categorizing. In other words, there may be no such thing as an "all-purpose" category representation.

These observations generate the prediction that different kinds of explanations should differentially impact categorization. Several accounts of explanation propose that explanations isolate the information likely to support future prediction and intervention, where this can be understood as a kind of inferential utility. One is likely to provide mechanistic explanations when generalizations of the kind captured by a physical stance are warranted, and functional explanations when generalizations of the kind captured by a design stance are warranted (see Lombrozo & Carey, 2006). The different ways of reasoning reflected by mechanistic and functional explanations may thus provide two ways to specialize multi-purpose category representations in the interest of particular inferences. In explaining a tove's slithiness by appeal to diet, one privileges diet as a feature and causal mechanisms as a basis for generalization. In explaining a tove's diet by appeal to gimbling, one privileges gimbling as a feature and functions as a basis for generalization.

The experiments below test the hypothesis that explanations influence categorization by examining whether functional explanations reverse previously documented effects of casual beliefs in determining the relative importance of features in categorization, called feature centrality (Sloman, Love, & Ahn, 1998). Specifically, Ahn and collaborators have documented a "causal status effect," according to which features that appear earlier in a causal chain are more central (Ahn, 1998; Ahn & Kim, 2000; see also Rehder, 2003 and Rogers & McClelland, 2004 for critical discussion). If a tove's diet causes slithiness, then a creature with a tove's diet but without a tove's slithiness should be judged more likely to be a tove than a creature with a tove's slithiness but without a tove's diet. In these experiments, 'causal status' corresponds to the explanatory privilege conferred by mechanistic explanations.

But if functional explanations reflect an alternative way to reason about causal structure, and this alternative supports different generalizations, then reasoning "functionally" may alter categorization judgments. In particular, explaining a feature by appeal to its effects may render the feature's functional affordances more important than proximate causes. More concretely, if one explains slithiness by appeal to gimbling, one may judge a creature with a tove's diet but without a tove's slithiness *less* likely to be a tove than a creature with a tove's slithiness but without a tove's diet.

To test these predictions, Experiment 1 examines whether participants who spontaneously explain a feature functionally rather than mechanistically are less likely to exhibit a causal status effect, while Experiment 2 examines whether prompting participants to provide a functional explanation eliminates effects of causal status.

2. Experiment 1

2.1. Participants

Ninety-six Berkeley students (66% female, mean age 20) completed the study in exchange for course credit.

2.2. Materials and procedures

The study consisted of a one-page questionnaire with a short paragraph introducing a novel category followed by a series of questions. Below is a sample item, with the questions labeled in italics:

> There is a kind of flower called a holing. Holings typically have brom compounds in their stems and they typically bend over as they grow. Scientists have discovered that having brom compounds in their stems is what usually causes holings to bend over as they grow. By bending over, the holing's pollen can brush against the fur of field mice, and spread to neighboring areas.

Explanation prompt: **Why do holings typically bend over**?

> Suppose you come across the following two flowers:
> Flower A bends over, but doesn't have brom compounds in its stem.
> Flower B has brom compounds in its stem, but doesn't bend over.

Categorization judgment: **Which flower do you think is more likely to be a holing?**
Circle one: Flower A / Flower B

F_1 item probability, $P(F_1)$: **How likely do you think it is that Flower A is a holing?**
Enter a probability between 0 and 100: __________

F_2 item probability, $P(F_2)$: **How likely do you think it is that Flower B is a holing?**
Enter a probability between 0 and 100: __________

F_2 conditional probability, $P(F_2|F_1)$: **Suppose a flower has brom compounds in its stem. How likely do you think it is that it bends over?**
Enter a probability between 0 and 100: __________

F_1 conditional probability, $P(F_1|F_2)$: **Suppose a flower bends over. How likely do you think it is that it has brom compounds in its stem?**
Enter a probability between 0 and 100: __________

Each novel category involved an item with two features, F_1 (e.g. brom compounds) and F_2 (e.g. bending over), where F_1 usually causes F_2 and F_2 serves a function (e.g. spreading pollen). There were a total of eight distinct categories, including four natural kinds and four artifacts. In addition to the holing, the natural kinds included an animal with red fur, a plant with irregular coloration, and an animal with blue feathers. The four artifacts were a mug with a removable handle, a refrigerator that turns off in cold weather, a ball that changes color, and a cup that releases water in low humidity. The categories were constructed to be plausible but sufficiently novel that participants would not have strong prior beliefs about the relative importance of features.

The order of the F_1 and F_2 probability questions was counterbalanced, as was the order of the conditional probability questions. Participants were randomly assigned to a questionnaire involving one of the two domains (natural kind, artifact), one of the four items within that domain, and one of the four possible question orders.

2.3. Results and discussion

Two coders classified participants' responses to the explanation prompt into one of three categories: those that only mentioned the cause (F_1), called *mechanistic*, those that mentioned the function, called *functional*, and those that did neither. Coder agreement was 100%. Overall, 67.7% of explanations were mechanistic, 21.2% were functional, and 1% were uncodable. The proportion of mechanistic explanations did not vary as a function of domain ($\chi^2(1) = 1.19$, $p = 0.28$), nor did the proportion of functional explanations ($\chi^2(1) = 1.41$, $p = 0.24$). However, among participants who provided functional explanations, 83% of those in the natural kind condition additionally mentioned the cause (F_1) while only 17% of those in the artifact condition did so ($\chi^2(1) = 13.03$, $p < 0.01$).

If explanations reflect or influence the perceived importance of category features, then categorization judgments should differ as a function of the kind of explanation a participant generated (see Table 1). Specifically, participants who generated functional explanations should be less likely than those who generated mechanistic explanations to privilege the item with F_1 but not F_2 over the item with F_2 but not F_1. For the categorization judgment, participants who generated a functional explanation were less likely to identify the item with F_1 as a category member (61% versus 74%), but this difference was not reliable ($\chi^2(1) = 1.85$, $p = 0.17$). As an alternative measure, a difference score consisting of the F_1 item probability minus the F_2 item probability was calculated for each participant [$P(F_1)$-$P(F_2)$]. Thus positive values reflect a belief that F_1 is more critical for categorization than F_2, and negative values the reverse. An ANOVA with explanation (mechanistic, functional) and domain (natural kind, artifact) as between-subjects factors and difference score as a dependent measure yielded a significant effect of explanation ($F(1,91) = 11.82$, $p < 0.01$), with no effect of domain ($F(1,91) = 0.65$, $p = 0.42$) nor an interaction ($F(1,91) = 0.02$, $p = .89$; see Fig. 1). As a group, participants who provided a mechanistic explanation judged the item with F_1 14% more likely to be a category member than the item with F_2, while those who provided a functional explanation judged the item with F_1 15% *less* likely to be a category member than the item with F_2 ($t(93) = 3.39$, $p < 0.01$, $r = 0.33$).

If a feature's role in explanation is important because it tracks inferential utility, then different explanations should correspond to different conditional probability judgments. To examine this, a difference score consisting of the F_2 conditional probability minus the F_1 conditional probability was calculated for each participant [$P(F_2|F_1)$-$P(F_1|F_2)$]. Thus a positive value reflects the belief that knowing about F_1 is more informative about the presence of F_2 than F_2 is about F_1, and negative values the reverse. Mirroring the findings with item probabilities, an ANOVA with explanation (mechanistic, functional) and domain (natural kind, artifact) as between-subjects factors and conditional probability difference score as a dependent measure yielded a significant effect of explanation ($F(1,91) = 6.02$, $p < 0.05$), with no effect of domain ($F(1,91) = 1.49$, $p = 0.23$) and a marginal interaction ($F(1,91) = 3.71$, $p = 0.06$; see Table 1). As a group, participants who provided mechanistic explanations judged the presence of F_1 10% more likely to establish the presence of F_2 than the other way around, while those who provided

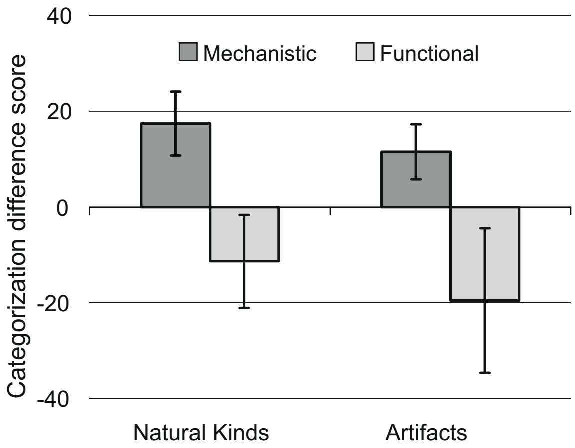

Fig. 1. Data from Experiment 1 as a function of domain and explanation classification. *Categorization difference score* is the average estimated probability that the item with the cause feature (F_1) is a category member minus the estimated probability that the item with the effect feature (F_2) is a category member. Error bars correspond to the standard error of the mean.

Table 1
Data from Experiment 1 as a function of domain and explanation classification. Data from the single participant with an uncodable explanation is excluded. The first data column (%F_1) reports the percent of participants who chose the item with feature F_1 over the item with feature F_2 in their categorization judgment. Means are followed in parentheses by the corresponding standard deviations.

Domain	Explanation	%F_1	$P(F_1)$	$P(F_2)$	$P(F_1)$ - $P(F_2)$	$P(F_2\|F_1)$	$P(F_1\|F_2)$	$P(F_2\|F_1)$ - $P(F_1\|F_2)$
Natural kinds	Mechanistic ($N = 30$)	63%	56 (27)	39 (31)	17 (36)	51 (34)	43 (31)	8 (24)
	Functional ($N = 18$)	50%	41 (22)	53 (25)	−11 (41)	50 (29)	44 (24)	5 (28)
Artifacts	Mechanistic ($N = 35$)	60%	56 (26)	45 (26)	12 (34)	76 (24)	64 (25)	12 (18)
	Functional ($N = 12$)	42%	43 (33)	63 (4)	−20 (52)	53 (31)	65 (21)	−12 (33)

functional explanations judged the presence of F_1 2% *less* likely to establish the presence of F_2 than the other way around ($t(93) = 2.17$, $p < 0.05$, $r = 0.22$).

These findings demonstrate that explanations are systematically related to a feature's importance in categorization, and are consistent with the stronger claim that explanations influence feature importance. However, a causal claim requires more than a correlational finding. Experiment 2 experimentally manipulates whether participants generate a mechanistic or a functional explanation and examines effects on categorization.

3. Experiment 2

3.1. Participants

One-hundred-ninety-two Berkeley students (61% female, mean age 21) completed the study in exchange for course credit. Three participants were replaced for failing to complete the study.

3.2. Materials and procedures

The study consisted of a questionnaire like that in Experiment 1, but the introductory paragraph did not specify the function of F_2. As in Experiment 1, participants in a *mechanism* condition and a *function* condition were asked to explain why members of the category have feature F_2, which is ambiguous as a request for a mechanistic or functional explanation. However, for participants in the *function* condition this question was immediately followed with: "What purpose might F_2 serve?" These participants were thus encouraged to answer the why-question with a functional explanation, but like participants in the *mechanism* condition, they were not told that F_2 serves a function nor what the function could be. Because participants never knew about functions with certainty, this manipulation was weaker than that in Experiment 1.

Items and counterbalancing were identical to Experiment 1. Participants were randomly assigned to questionnaires from the *mechanism* or *function* conditions.

3.3. Results and discussion

As in Experiment 1, two coders classified participants' responses to the explanation prompt as *mechanistic*, *functional*, and other. Coder agreement was 98%. In the *mechanism* condition, 97% of explanations were mechanistic, 1% were functional, and the remaining 2% were neither. In the *function* condition, 19% were mechanistic and 81% were functional. Participants were significantly more likely to provide mechanistic explanations in the *mechanism* condition than in the *function* condition ($\chi^2(1) = 120.12$, $p < 0.01$), and to provide functional explanations in the *function* condition than in the *mechanism* condition ($\chi^2(1) = 127.52$, $p < 0.01$), confirming that the condition manipulation had the intended effect. As in Experiment 1, there were no domain differences in the proportion of mechanistic explanations ($\chi^2(1) = 0.02$, $p = 0.88$) or functional explanations ($\chi^2(1) = 0.02$, $p = 0.88$), and a large proportion of participants who provided a functional explanation also mentioned feature F_1 (84%). However, this proportion did not vary as a function of domain (83% for natural kinds versus 85% for artifacts, $\chi^2(1) = 0.06$, $p = 0.80$).

If explanations do not merely reflect different judgments concerning the importance of category features, but also have a causal impact on such judgments, then categorization judgments should differ as a function of condition (see Table 2). Specifically, participants in the *function* condition should be less likely than those in the *mechanism* condition to privilege the item with F_1 but not F_2 over the item with F_2 but not F_1. For the categorization judgment, participants in the function condition were significantly less likely to identify the item with F_1 as a category member (55% versus 71%; $\chi^2(1) = 5.03$, $p < 0.05$). As an alternative measure, a categorization difference score was calculated as in Experiment 1. An ANOVA with condition (*mechanism*, *function*) and domain (natural kind, artifact) as between-subjects factors and difference score as a dependent measure yielded a significant effect of condition ($F(1,188) = 8.27$, $p < 0.01$), with no effect of domain ($F(1,188) < 0.01$, $p = 0.97$) nor an interaction ($F(1,188) = 2.74$, $p = 0.10$; see Fig. 2). As a group, participants in the *mechanism* condition judged the item with F_1 15% more likely to be a category member than the item with F_2, while those in the *function* condition judged the item with F_1 1% *less* likely to be a category member than the item with F_2 ($t(186) = 2.87$, equal variances not assumed, $p < 0.01$, $r = 0.21$).

To examine whether the *mechanism* and *function* conditions yielded different conditional probability judgments, a conditional probability difference score was calculated as in Experiment 1. Once again mirroring the findings with item probabilities, an ANOVA with condition (*mechanism*, *function*) and domain (natural kind, artifact) as between-subjects factors and conditional probability difference score as a dependent measure yielded a significant effect of condition ($F(1,188) = 5.97$, $p < 0.05$), with no effect of domain ($F(1,188) = 0.233$, $p = 0.63$) nor an interaction ($F(1,188) = 0.73$, $p = 0.39$; see Table 2). As a group, participants in the *mechanism* condition judged the presence of

Table 2
Data from Experiment 2 as a function of domain and condition. The first data column (%F_1) reports the percent of participants who chose the item with feature F_1 over the item with feature F_2 in their categorization judgment. Means are followed in parentheses by the corresponding standard deviations.

Domain	Condition	%F_1	$P(F_1)$	$P(F_2)$	$P(F_1)$ - $P(F_2)$	$P(F_2\|F_1)$	$P(F_1\|F_2)$	$P(F_2\|F_1)$ - $P(F_1\|F_2)$
Natural kinds	Mechanism ($N = 48$)	69%	51 (25)	41 (24)	10 (40)	49 (29)	38 (24)	11 (28)
	Function ($N = 48$)	65%	50 (22)	47 (21)	4 (34)	47 (25)	42 (24)	5 (22)
Artifacts	Mechanism ($N = 48$)	73%	60 (23)	41 (24)	19 (41)	73 (19)	57 (24)	16 (29)
	Function ($N = 48$)	46%	51 (21)	56 (22)	−6 (37)	70 (22)	66 (22)	4 (26)

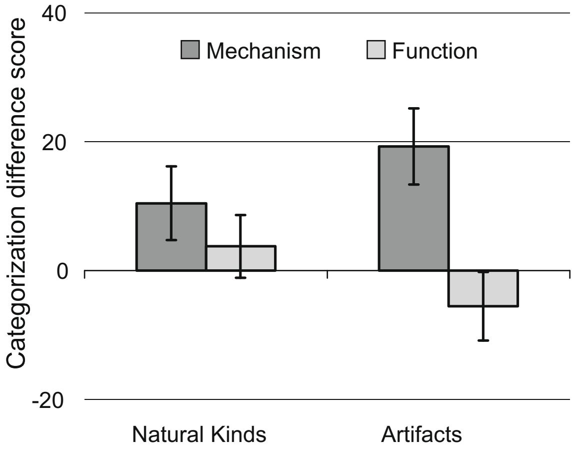

Fig. 2. Data from Experiment 2 as a function of domain and condition. *Categorization difference score* is the average estimated probability that the item with the cause feature (F_1) is a category member minus the estimated probability that the item with the effect feature (F_2) is a category member. Error bars correspond to the standard error of the mean.

F_1 14% more likely to establish the presence of F_2 than the other way around, while those in the *function* condition judged the presence of F_1 only 5% more likely to establish the presence of F_2 than the other way around ($t(184) = 2.45$, equal variances not assumed, $p < 0.05$, $r = 0.18$).

4. General discussion

Experiment 1 found a correlation between how a property was explained and subsequent categorization judgments, consistent with the hypothesis that different kinds of explanations reflect differences in underlying reasoning with consequences for categorization. In particular, participants who generated mechanistic explanations exhibited a causal status effect, while those who generated functional explanations did not. Experiment 2 found that prompting participants to explain either mechanistically or functionally had similar effects, consistent with the stronger claim that explanations causally influence the perceived importance of features in categorization. Moreover, both experiments found that differences in categorization judgments tracked differences in the perceived inferential utility of features, as assessed by conditional probability judgments.

These findings contribute to a growing body of work suggesting that explanation and categorization are intimately related (Keil, 2006; Lombrozo, 2006; Murphy, 2002). Murphy and Medin (1985) proposed that concepts are coherent by virtue of the theories in which they are embedded, where theories are "any of a host of mental 'explanations'" (see also Carey, 1985; Gopnik & Meltzoff, 1997; Keil, 1989; Rips, 1989). Empirical work supports this proposal (e.g. Ahn, Marsh, Luhmann, & Lee, 2002; Murphy & Allopenna, 1994; Patalano, Chin-Parker, & Ross, 2006; Rehder & Hastie, 2001; Wisniewski, 1995), and the current findings further indicate that explanations may influence the very representations or processes involved in categorization.

These experiments do not address the mechanisms by which explanations can influence categorization, but two possibilities are worth distinguishing. First, one possibility is that explanations can play an active role in structuring conceptual representations. By identifying meaningful relationships, explanations could generate a structured representation that serves as input to the mechanisms involved in categorization. This possibility is consistent with research concerning the instability of conceptual representations and recognizing a role for representations that appear to be constructed 'on-the-fly' (e.g. Barsalou, 1987). A less radical possibility is that explanations serve as a cue to the information-rich clusters category representations are intended to track. Specifically, if explanations highlight inferentially useful information, their content may be an effective guide to underlying structure in the world. On this view, explanations should exert an especially large role when prior beliefs are minimal, as with artificial categories like Carroll's slithy toves, because alternative cues to underlying structure are less likely to be available.

If functional explanations are understood causally (and there's evidence that they are, see Lombrozo & Carey, 2006), then situations that warrant functional explanations may just be those for which functional information has a "deeper" causal status than the proximate causes cited by mechanistic explanations. Indeed, Ahn (1998) attempts to assimilate effects of functional information to causal status, and causal status is itself partially motivated in terms of inferential utility (Ahn, 1998; Proctor & Ahn, 2007). Ahn (1998) reports a correlation between the "because" statements participants endorse and the importance of functions in categorization decisions. For example, participants who give higher ratings to claims involving functions as causes, such as "mirrors are made of glass because they reflect an image," than to claims involving functions as effects, such as "mirrors reflect an image because they are made of glass," are more likely to privilege functions in categorization.

However, the current proposal differs from causal status, even broadly understood. The sense in which being made of glass *causes* a mirror to reflect an image is quite different from the sense in which reflecting an image *causes* a mirror to be made of glass. These two notions of 'cause' reflect the different dependence relationships captured by mechanistic and functional explanations. Rather than regarding one causal relation as more basic, one can extend the insight from the literature on stances, noting that the very same causal system can support multiple generalizations. Depending on the judgment being made, a different notion of 'cause' may support the relevant generalizations. The distinction between mechanistic and functional explanation provides a natural way in which to understand this flexibility.

Additional evidence for the utility of invoking explanatory rather than exclusively causal concepts would come from the finding that non-causal explanations impact the importance of features in categorization (see Prasada & Dillingham, 2006, for indirect but suggestive evidence that this is so), or that the quality of explanations modulates effects of causal status (Jameson & Gentner, 2008). For example, simple mechanistic explanations are typically preferred over complex alternatives (Lombrozo, 2007), and should thus generate correspondingly larger effects of causal status. These and other questions await future work.

Acknowledgements

Thanks to Christina Botros, Minh-Chau Do, Beibei Luo, Lisa Qin, and Emily Wei for help with data collection. A subset of these results were presented at the annual meetings of the Society for Philosophy of Psychology (June, 2007) and the Cognitive Science Society (August, 2007). This work was partially supported by an NSF grant (BCS-0819231) awarded to the author.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? *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* (Vol. 40, pp. 23–65). New York: Academic Press.

Ahn, W., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory and Cognition, 30*, 107–118.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge: Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carroll, Lewis. (1871/2008). Through the Looking Glass. Retrieved October 1, 2008. Available from <http://www.gutenberg.org/files/12/12-h/12-h.htm>.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Jameson, J., & Gentner, D. (2008). Causal status and explanatory goodness in categorization. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *In Proceedings of the 30th annual conference of the cognitive science society* (pp. 291–296). Austin, TX: Cognitive Science Society.

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

Keil, F. C. (1994). The birth and nurturance of concepts by domains: the origins of concepts of living things. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind* (pp. 234–254). Cambridge: Cambridge University Press.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Science, 10*, 464–470.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232–257.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.

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

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory and Language, 54*, 407–424.

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73–112.

Proctor, C., & Ahn, W. (2007). The effect of causal knowledge on judgments of the likelihood of unknown features. *Psychonomic Bulletin and Review, 14*, 635–639.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: the effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*, 323–360.

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

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rosch, E. (1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–206). Cambridge, MA: MIT Press.

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

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