Title
The Inductive Benefit of Being Far Out: How Spatial Location of Evidence Impacts Diversity-based Reasoning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Lawson, Chris A.
Wolfe, Noah

Publication Date
2019

Peer reviewed

The Inductive Benefit of Being Far Out: How Spatial Location of Evidence Impacts Diversity-based Reasoning

Chris A. Lawson (lawson2@uwm.edu)
Noah Wolfe (wolfena@uwm.edu)
University of Wisconsin – Milwaukee
Department of Educational Psychology
Milwaukee, WI USA

Abstract

Inductive reasoning is constrained by several principles that govern how we choose to generalize evidence to new cases. Here we focus on diversity principle of induction, which describes the tendency to favor inductive arguments that include a diverse sample of evidence over those that include a homogenous sample of evidence. Several studies reveal that adherence to the diversity principle is influenced by a range of conceptual processes, such as an individuals' prior knowledge or expectations about the categories and properties represented in the evidence. In the two experiments reported here we examined a contextual factor of the available evidence – the spatial separation of evidence exemplars – that we expected would impact how people reason about diverse samples. We found that when the pictures (Experiment 1) or labels (Experiment 2) used to represent evidence exemplars were presented far apart (approximately 10 cm), participants showed a greater willingness to endorse arguments with diverse exemplars than those with homogenous sample, relative to when these exemplars were placed in close proximity (approximately 1 cm apart). We discuss these results as they relate to existing models of induction.

Keywords: Inductive reasoning; Generalization; Diversity principle; Situated cognition

Introduction

Inductive reasoning, the process by which we use specific facts to arrive at general conclusions, is critical to our cognitive lives. For example, learning that hawks have hollow bones serves as evidence to support the inductive inference that other birds are likely to have hollow bones. Given the powerful role of induction for a range of cognitive processes there has been considerable interest in determining the constraints that guide the inferences we make. For example, in their classic work, Osherson and colleagues (1990) outlined several inductive principles that systematically constrain how we use evidence to arrive at inductive decisions. The present study focused on one such principle – the diversity principle of induction. Consider the two arguments below in which the two statements above the lines represent evidence and the statement below the lines represents a conclusion:

Hawks have hollow bones
Penguins have hollow bones (1)
Larks have hollow bones

Hawks have hollow bones
Eagles have hollow bones (2)
Larks have hollow bones

When asked to judge which of these two represent stronger inductive arguments, participants tend to select those that include a diverse sample of exemplars (1) rather than those that include a homogenous sample of exemplars (2) (Heit, Hayes, & Feeney, 2005; Kim & Keil, 2003; Osherson, et al., 1990; also, Lopez, 1995).

Most explanations of diversity effects focus on the ways individuals represent the content (i.e., categories and to-be-generalized properties) of the available evidence. For example, Osherson et al. proposed the similarity-coverage model to account for diversity effects. This perspective posits that individuals first consider the overarching category about which the inductive judgment should be considered. In the two inductive arguments presented above the coverage category in *bird*. Participants then assess the extent to which the evidence in each set of arguments covers this overarching category. According to this model individuals rely on their calculation of the similarity between exemplars to assess the extent to which each sample covers the conclusion category. Greater dissimilarity between exemplars within the evidence sample reflects greater coverage of the category, and therefore facilitates diverse-based reasoning.

Diversity effects have also been explained as Bayesian inference. From this perspective individuals rely on their prior beliefs about categories and properties to test hypothesis about the scope of property projection (Heit, 1998; Lo, Sides, Rozelle, & Osherson, 2002). Our prior experience may lead us to believe that some categories (e.g., hawks and eagles) share many features in common and others categories (e.g., hawks and penguins) share fewer features. Thus, we are not surprised to learn about a new property that happens to be shared by two categories we have heretofore expected share many properties. In contrast, we are surprised to learn about a property that is shared by two categories that we believed had very little in common. This surprising sample of evidence, coupled with our expectation that samples of evidence tend to be selected purposefully (Lawson & Kalish, 2009), makes the diverse sample a better argument to support a conclusion about a superordinate category.

There are notable cases in which individuals fail to adhere to the diversity principle. For example, several studies have shown that individuals with rich domain knowledge are less likely to consider taxonomic diversity in lieu of other evidence (Lopez, Atran, Coley, Medin, & Smith, 1997; Proffitt, Coley, & Medin, 2000). Moreover, when experts do engage in diversity-based reasoning they rely on a range of strategies that appeal to their knowledge about the domain, such as the types of properties that are transmitted across categories, or the relative size of the category that is represented in the samples of evidence (Proffitt et al., 2000). Thus, experts will depart from using taxonomic diversity as a basis for induction under conditions in which their rich domain knowledge suggests an alternative inductive strategy is optimal.

In related work Medin and colleagues (2003) showed that non-experts (college students) prefer to generalize from a sample of evidence that highlights a relevant relation between two evidence exemplars rather than a sample that includes taxonomically diverse exemplars. For example, participants judged an argument in which fleas and butterflies were attributed the same property as better support to conclude that the property is true of sparrows, than an argument in which fleas and dogs were the same property. This latter sample signals a relevant causal relationship that draws attention away the greater taxonomic diversity of the two exemplars, thereby leading individuals to favor the inductive argument with less diverse exemplars.

These exceptions are notable for two reasons. First, they highlight the role of prior knowledge about categories and properties when reasoning about the content of an inductive problem. Second, they bring to light an important methodological point: specific task modifications, such as the type of property or the relationship between categories presented in the evidence, impact how people reason about diverse samples. In support of this point, Feeney and Heit (2011) showed that the content of the to-be-generalized property serves as a prime to either encourage or discourage diversity-based responses. In their study participants exhibited diversity effects when they were primed with a general property that can be construed as common across a wide range of category members (e.g., are warm-blooded), but did not show these effects when they were primed with an idiosyncratic property (e.g., lives in the water).

In the present studies we examined whether contextual factors, such as how evidence exemplars are presented, may impact the extent to which participants obey the diversity principle. We were particularly interested in the potential influence of the spatial location of exemplars for two related reasons. The first concerns findings from research demonstrating that taxonomic categories tend to be, in many ways, represented within a multidimensional space which can be described as reflecting psychological distance between exemplars (Collins & Quillian, 1969; Hutchinson & Lockhead, 1977; Rips, 1975; Schaeffer & Wallace, 1969). Among other things this psychological distance can be created by similarity relations; for example, relative to their membership within the bird category robins and sparrows can be considered *close* (they share many properties) whereas robins and ostriches are *far* apart (they share few properties). From this perspective, diverse samples are likely to be those that represent items that have greater representational distance.

The second, related, idea comes from research on situated cognition and embodiment (Barsalou, 2006; Wilson, 2002), in which it has been argued that the way we think about and represent concepts is determined, at least in part, by the way we experience and engage with concepts. For example, in addition to activating semantic features, many of the concepts we reason about (e.g., dogs) activate motor and sensory features (e.g., throwing to-be-retrieved items, going for walks, tugging on a leash, etc.) that reflect simulations of how we might interact with concepts (Barsalou, 2006). At a broad level, the embodiment framework challenges cognitive models to consider the role of the environment for a cognitive system (e.g., Hutchins, 1995).

With theses issues in mind we examined whether creating greater physical distance between exemplars within a sample would impact diversity-based reasoning. In two experiments participants were given inductive problems in which three evidence exemplars were presented either in close proximity to each other (within 1 cm), or far from each other (approximately 10 cm apart) (See Figure 1). Half the evidence samples included a diverse range of exemplars and the other half included a homogenous set of exemplars. Our main prediction was that the greater separation of items would encourage participants to consider the coverage, or range, of the exemplars and therefore would lead to higher ratings for inductive arguments that included diverse samples compared to conditions in which the items were spaced close together.

The experiments assessed two additional factors. The first concerns the contents of the evidence samples. In Experiment 1 the items were represented by pictures of animals used to represent the categories presented in the evidence, whereas in Experiment 2 the items were represented by category labels (see Figure 1). This manipulation allowed us to test whether any potential effects of evidence spacing were due to perceptual processed that governed the way participants compared the physical features of the exemplars (i.e., differences

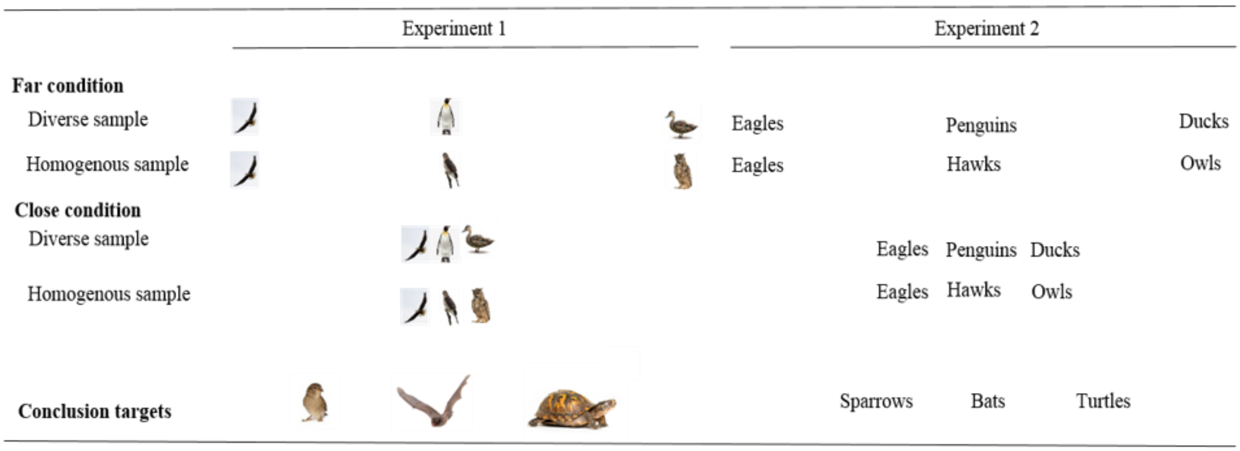

Figure 1. Schematic representation of the design of Experiments 1 and 2.

between the pictured items), rather than differences in how they represented the categories.

Finally, we asked participants to make inductive judgments about three different targets, each of which varied in similarity to/taxonomic distance from the category represented by the evidence exemplars. For example, for the item in which evidence exemplars were represented by birds, participants were asked to make a judgement about a new bird (e.g., sparrow), a bat, and a reptile (e.g., turtle). We expected that the spacing effects would be limited to the category that represents the lowest level of abstraction covered by the evidence and target exemplars (e.g., Osherson et al., 1990). However, if the spatial manipulation has a more general effect on how individuals compare stimuli, it is possible the spaced presentation could lead to an overall increase in one's willingness to generalize from diverse samples to any targets. Thus, this manipulation allowed us determine the extent to which varying the spatial location of the evidence influenced participants' adherence to the diversity principle, in particular.

Experiment 1

Participants. Fifty-three undergraduate students participated for extra credit in a college course. Participants were recruited from, and were representative of, a medium-sized Midwestern US city.

Design. This experiment employed a 2 x 2 x 3 design with Spatial location of evidence exemplars (Close, Far) manipulated between subjects and Sample composition (Diverse, Non-diverse) and Conclusion target (Same basic-level, Similar superordinate, Dissimilar superordinate) manipulated within subjects. Participants were randomly assigned to either the Close condition or the Far condition such that there was an approximately equal number of participants in the two conditions: Close (N = 27), and Far (N = 26).

Materials. Participants were presented 12 inductive reasoning problems each of which included a sample comprised of 3 evidence exemplars. Half of the samples included a diverse set of exemplars (e.g., eagles, penguins, ducks) and the other half included a homogenous set of exemplars (e.g., eagles, hawks, owls). A novel biological property (e.g., Enzyme A) was attributed to the exemplars within the sample. A different novel property for each of the twelve problems.

For each reasoning problem participants were asked to make judgements for 3 different conclusion targets each of which varied in relation (taxonomic and/or perceptual relatedness) to the category covered by the evidence exemplars (see Figure 1 for a sample item). One target was drawn from same basic-level category that was represented by the evidence exemplars (e.g., sparrows). The other two target categories were drawn from superordinate categories (e.g., bats and turtles). Each of the evidence exemplars and targets were represented by photographs of a single animal (2cm x 2cm).

Procedure. The experiment was conducted on a desktop computer with a 24" screen. The spatial arrangement manipulation involved varying the location of the evidence exemplars as they appeared on the screen. In both conditions the exemplars were presented in row on the top of the screen. In the *Far* condition the exemplars were spaced so that there was an approximately 10 cm gap between each. In the *Close* condition the exemplars were bunched together so that there was an approximately 1 cm gap between each. For each item the three evidence exemplars were presented at the same time and were accompanied by a statement (appearing below three exemplars) that attributed a property to all the animals (e.g., "these animals have Enzyme A").

After the evidence exemplars were presented participants were asked to make a judgment about each of the three conclusion targets. A photograph of an animal used to represent the category was presented approximately 6 cm below the evidence exemplars and was accompanied by a prompt to judge the likelihood that the target category would have the property that was attributed to the evidence exemplars (e.g., "How likely is it that sparrows have Enzyme A?"). Participants were instructed to use a scale, ranging from 0 ("not at likely" to 100 ("very likely), to determine their likelihood judgment. The three target categories were presented in random order.

Note that because sample diversity was manipulated within subjects we counterbalanced across participants which category was represented by a diverse sample or homogenous sample of exemplars. Also the order of presentation of diverse and homogenous samples (within participants) was randomized.

Results

Average likelihood ratings were submitted to a mixed ANOVA with Spatial arrangement of evidence (Close, Far) as the between subjects variable and Sample composition (Diverse, Homogenous), and Conclusion target as the within subjects variables. The only significant main effect was Conclusion target, $F(2, 102)=374.93$, $p<.001$, $\eta^2>.87$, due to a stepwise decrease in likelihood ratings as a function of the decrease in similarity/increase in taxonomic distance from the evidence exemplars to the target categories, all *p*s<.001 Tukey's HSD.

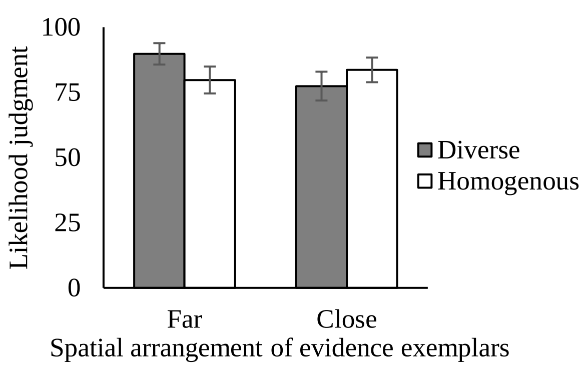

Figure 2. Mean likelihood judgements for basic-level targets for Diverse samples and Homogenous samples in both Spatial arrangement conditions in Experiment 1. Bars represent 1+/- SE from the mean.

The analysis also yielded several noteworthy interactions, all of which were captured by a 3-way interaction, $F(2,102)=4.72$, $p<.02$, $\eta^2=.09$. Analyses of the effects of spatial location on the ratings for each target revealed different patterns of responses for only the Basic-level targets. Simple effects analyses indicated there was a sample location by sample diversity interaction for basic-level targets $F(1,51)=8.49$, $p=.005$, $\eta^2=.17$. As suggested by Figure 2, this interaction was due to differences in responses for Diverse samples of evidence, for which the ratings were significantly higher in the Far condition than the Close condition, $p<.001$. Additional comparisons indicated that were significant differences in ratings between diverse samples and homogenous samples in the Far condition $F(1,25)=7.89$, $p=.01$, $\eta^2=.12$, but not in the Close condition ($F<1.50$). No other effects or interactions were significant (all *F*s<1.60).

Discussion

These results indicate that the spatial location of evidence exemplars had a consistent and precise effect on judgments about diverse, but not homogenous, samples. Adults consistently gave higher likelihood ratings for diverse samples when the evidence exemplars were separated from each other than when they were presented in close proximity to each other. However, the effect of spatial location was only present for targets from the same, basic-level, conclusion category as the evidence exemplars. Thus, these results provide preliminary evidence in support of our prediction that contextual factors, such as the way evidence is presented, can facilitate diversity-based reasoning.

Experiment 2

This experiment was designed to address at least two concerns raised by Experiment 1. First, because the items were represented by a photograph of a single animal it remains unclear if participants interpreted the exemplars as representative of the categories they were intended to represent or if they interpreted the evidence as representative of single individual concepts. Second, because the materials included photographs it is possible the effects were due to differences in how participants compared the stimuli, rather than their assessment of the diversity represented by the categories in the evidence. We addressed both of these concerns in this experiment by replacing the photographs with category labels.

Method

Participants. Forty-nine undergraduate students participated for extra credit. Participants were recruited from, and representative of the population of, a medium-sized Midwestern US city.

Design, Materials, and Procedures. This experiment was identical to Experiment 1 in every respect except the stimuli. In this case, rather than presenting photographs to represent the evidence and target items, participants were presented category labels. Participants were randomly assigned to the Far ($N=24$) or Close ($N=25$)

conditions. See Figure 1 for a schematic of the study design.

Results

The analysis replicated the pattern of results that was found in Experiment 1. Again there was a three-way interaction between Spatial arrangement, Sample Composition, and Conclusion target, $F(2, 92)=3.48$, $p=.02$, $\eta^2=.078$. Further analysis revealed a significant Sample composition by location effect interaction for Basic-level targets, $F(1, 46) = 10.32$, $p=.002$, n=.18. As suggested by Figure 3, the interaction was due to higher ratings for Basic-level targets for diverse samples than homogenous samples in the Far condition, $F(1, 47) = 12.12$, $p<.001$, $\eta^2=.13$, but not in the Close condition, $F<1.00$. Also, participants exhibited higher likelihood ratings for diverse samples in the Far condition than in the Close condition, $F(1,97)=6.42$, $p=.03$, $\eta^2.09$ As was the case in Experiment 1, no other effects or interactions were significant.

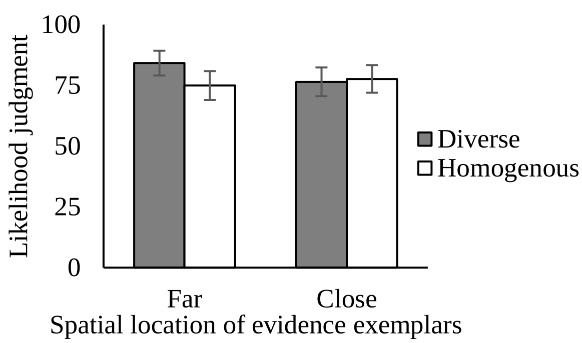

Figure 4. Mean likelihood judgements for basic-level targets for Diverse samples and Homogenous samples in both Spatial location conditions in Experiment 2. Bars represent 1+/- SE from the mean.

Discussion

These results replicated those found in Experiment 1. Participants rated arguments with diverse samples as providing better support for conclusions than arguments with homogenous samples when the exemplars were spatially distant from each other but not when these same evidence exemplars were in close proximity. Moreover, the diverse sample were given higher ratings when the evidence exemplars were more distant than when they were close. These results suggest that the findings from Experiment 1 were not due to participants' interpretation of the evidence as applying to specific individuals, rather than categories, nor their reliance on perceptual features of the task.

General Discussion

Prior research indicates that diversity-based reasoning is dictated by our knowledge or expectations about the categories, or the to-be-generalized properties that are represented in the evidence (Feeney & Heit, 2011; Heit, 1998; Osherson et al., 1990). Thus most existing models account for diversity effects by focusing on how people reason about the content of an inductive problem. In the two experiments reported here we demonstrated that certain contextual factors, such as the way exemplars are presented, also contribute to diversity-based reasoning. Specifically, participants showed a greater willingness to endorse arguments that included diverse samples when these samples were presented in such a way that there was a large spatial separation between each of the evidence exemplars relative to when the same exemplars were presented without a large separation between evidence exemplars. In other words, diversity effects were strongest when the evidence covered more physical space.

It is difficult to reconcile these results with current explanations for diversity effects. One could argue that the spacing effects in Experiment 1 are consistent with the feature-based induction model of induction (Sloman, 1993) insofar as the presentation impacted the way participants compared stimuli, or identified overlapping or unique features, and thus impacted their calculation of diversity. However, this interpretation does not account for the observed effects in Experiment 2 in which the stimuli were represented by labels rather than images. The results are also inconsistent with the idea that a calculation of similarity between the evidence exemplars is sufficient to assess category coverage (Osherson et al., 1990). The similarity coverage model does not account for the finding that participants gave higher ratings for diverse samples when the evidence exemplars were spread far apart compared to when they were positioned close together.

Those in favor of the Bayesian or Relevance accounts of induction might interpret the results as the outcome of pragmatic factors. It could be argued that participants assumed that the exemplars were purposefully placed in close proximity or far apart. According to the Relevance theory of induction (Medin et al., 2003), participants rely on standard rules of communication (e.g., Grice, 1975; Sperber & Wilson, 1995), such as the notion that people present information in such a way as to highlight a relevant piece of information. Thus, it could be argued that participants reasoned as-if the exemplars were deliberately placed apart to draw attention to the coverage of the exemplars (or placed together to highlight the similarities between them). Although these models can accommodate these effects of spatial location, they do not explain them. Assuming participants reasoned that exemplars were spread apart purposefully, why would they interpret this decision was intended to highlight the coverage of evidence exemplars?

Our interpretation of these results is that they provide some support for the grounded, or situated, aspect of diversity-based reasoning. Most models of situated cognition tend to focus on the influence of different modalities (e.g., motor sequences) on conceptualization. Diversity refers to a feature of samples, not isolated concepts. Diverse samples are those that include exemplars that provide coverage of a category; diverse samples occupy greater psychological space. Here we showed that presented evidence in such a way that the sample occupied greater physical space facilitated diversity-based reasoning. That the spacing effects were not observed for homogenous samples, or for targets from more distant conclusion categories, suggests that spacing did not influence *whether* participants interpreted samples as diverse. Rather, the results indicate that the broad spacing of exemplars primed cognitive processes that can draw attention to sample diversity, and thereby strengthen diversity effects.

Clearly, more work is needed to better understand the scope of these effects. For example, to clarify the potential impact of participants' pragmatic assumptions it will be important to determine whether we can replicate these effects in conditions in which participants are made to believe that the location of the exemplars was not chosen deliberately. Additionally, because participants did not show the diversity effect in the Close condition it will be important to replicate these findings with a different set of stimuli. Also, it will be important to explore other ways in which individuals might be primed to consider the breadth or scope of evidence. For example, we are currently exploring the impact of gestures on adults' and children's adherence to diversity and sample size principles of induction.

There are several notable limitations of these experiments. First, the methods were different from those typically used on the inductive reasoning literature. Participants are often given arguments and asked to determine the sample that provides the best support for a conclusion. Here the evidence exemplars were presented as single photographs or category labels, rather than premises in an inductive argument. The effect of spacing might have been pronounced because this method encouraged participants to compare the stimuli. Also, although the observed effects were consistent, they were rather small. It will be important to replicate these results with stimuli from different domains to be sure these effects are not exclusive to the set of items.

Despite these limitations, these results raise important questions about the impact of contextual factors on inductive reasoning. In particular we showed that presenting evidence in such a way that it covered a broad physical space provided greater support for diversity-based reasoning than when the same evidence was presented in a narrow physical space. While we do not deny that there is still much to learn about the influence of category knowledge, and prior beliefs, on inductive reasoning, these experiments call for more work on understanding the impact of contextual features on induction. As much as inductive reasoning is influenced by *what* is presented in an inductive problem, it seems intuitive that it would be influenced, at least to a certain degree, by *how* evidence is presented.

References

Barsalou, L.W. (2006). Situated conceptualization: Theory and application. In Y. Coello & M.H. Fischer (Eds.), *Foundations of embodied cognition.* East Sussex, UK: Psychology Press.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-248.

Feeney, A., & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning, 17*, 156-181.

Grice, H.P. (1975). "Logic and Conversation," In P. Cole and J. Morgan, *Syntax and Semantics*. Academic Press.

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

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N Chater (Eds.), Rational models of cognition (pp. 248-274). Oxford, UK: Oxford University Press.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin*. Washington, DC: American Psychological Association.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Hutchinson, J.W., & Lockhead, G.R. (1977). Similarity as distance: A structural principle for semantic memory. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 660-678.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition, 31*, 155-165.

Lawson, C.A., & Kalish, C.W. (2009). Sample selection and inductive generalization. *Memory & Cognition, 37,* 596-607.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26*, 181-206.

Lopez, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition, 23*, 374-382.

Medin, D. L., Coley, J.D., Storms, G., & Hayes, B.K. (2003). A relevance theory of induction.

Psychonomic Bulletin & Review, 10(3), 517-532.
Osherson, D.N., Smith, E.E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185-200.
Proffitt, J. B., Coley, J. D., Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 811-828.
Rips, L.J. (1975). Inductive judgements about natural categories. *Journal of verbal learning and verbal behavior*, *14*, 6.
Sloman, S.A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231-280.
Schaeffer, B., & Wallace, R. (1969). Semantic similarity and the comparison of word meanings. *Journal of Experimental Psychology*, *82*, 343-346.
Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition.* Wiley-Blackwell.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625-636.