6

Defending Diversity

Evan Heit, Ulrike Hahn, and Aidan Feeney

Why do some observations lead to broad generalizations whereas other observations do not have as much influence on people's beliefs? One principle of evaluating evidence is the diversity principle, which states that more diverse evidence should lead to stronger inferences than a narrow sample of evidence. For example, if you see someone repeatedly acting aggressively while giving lectures, you might infer that this behavior is simply the person's lecturing style and not draw broad inferences about the person in general. On the other hand, if you observe someone acting aggressively in diverse contexts, such as in a lecture, at a restaurant, and at a party, you might infer that this is indeed an aggressive person overall.

There are two views on the diversity principle. The first is the normative view. The emphasis in this view is on why people ought to reason in accord with the diversity principle—that is, why it will lead to successful reasoning. The normative view encompasses historical evidence of expert scientific practice following the diversity principle, as well as various arguments and proofs intended to show that the diversity principle leads to useful inferences. The second view is descriptive. The descriptive view emphasizes showing the many situations in which ordinary reasoning does indeed follow the diversity principle. Typically, the descriptive view has relied on experimental evidence investigating various subject populations. This chapter is organized in terms of these views. The first two sections briefly summarize the positive cases for the normative view and the descriptive view. The following two sections describe some of the challenges to these views and offer some responses in defense of the diversity principle.

Doug Medin's many contributions to this line of work and related research, inside and outside the lab, are gratefully and respectfully acknowledged.

The Normative View

The diversity principle has been historically important to philosophers of science in describing scientists' preference for testing a theory with a diverse set of experiments rather than repeatedly conducting the same experiment or close replications. An early example was in Bacon's *Novum Organum* (1620/1898), which cautioned scientists of the day against inferences drawn from narrow samples. Bacon illustrated this point with the concept of heat, listing 28 different kinds of heat and hot things that would need to be observed in a study of heat. In a more modern example, Salmon (1984) described how early in the 20th century, scientists had developed a wide variety of experimental methods for deriving Avogadro's number (6.02×10^{23}), the number of particles in a mole of any substance. These methods included Brownian movement, alpha particle decay, X-ray diffraction, black body radiation, and electrochemistry. Together, these techniques strongly supported the existence of atoms and molecules. Salmon argued that any one of these techniques taken alone, no matter how carefully applied, would not have been sufficient to persuade scientists of that period to accept the atomic theory over its principal rival, energeticism, which conceived of matter as being continuous rather than being composed of particles. It was the diversity of evidence that led to this major change in scientific belief.

This historical approach is complemented by attempts of philosophers and statisticians to argue for or even prove the benefits of following the diversity principle. For example, Nagel (1939) argued that to obtain more reliable estimates, a scientific theory should be derived from diverse observations rather than many similar observations. He used the example of inspecting the quality of coffee beans delivered on a ship, proposing that inspecting small samples of beans from various parts of the ship would be better than inspecting a large number of beans from just one location. Carnap (1950) linked the collection of diverse evidence to the desirable quality of scientific theories that theories should make novel predictions rather than merely redescribe old data. A scientific theory should be strongly supported if it makes diverse predictions that are subsequently supported. Similarly, Hempel (1966) related the collection of diverse evidence to a falsifying research strategy. Namely, it is better to test theories with a wide variety of challenging experiments rather than conducting a series of similar experiments that seem very likely from the outset to be successful.

These intuitions have led to several attempts to formalize the advantage for following the diversity principle. As reviewed by Wayne (1995), there have been two lines of approach. The first approach compares correlated sources of evidence to independent sources of evidence. For example, if a person is seen giving a lecture aggressively, observing another aggressive lecture from this person does not seem to add much independent or surprising information. In contrast, observing aggressive behavior by this person in a restaurant would seem less predictable because of the lower similarity between the two contexts. Hence, seeing the person act aggressively in diverse contexts provides stronger evidence to promote further inferences. For formal treatments of this correlation approach, linking similarity to probability theory, see Earman (1992) and Howson and Urbach (1993).

The second approach is the eliminative approach. The idea behind the eliminative approach is that diverse data sets will be particularly useful for eliminating plausible but incorrect hypotheses, allowing stronger inferences to be drawn on the basis of the remaining, contending hypotheses. In contrast, nondiverse data sets will likely be consistent with too many hypotheses to allow any strong inferences. For example, seeing someone act aggressively in a lecture, at a restaurant, and at a party will serve to eliminate the hypothesis that this person simply acts aggressively at lectures. For a formal treatment of this approach, including a geometric proof, see Horwich (1982), and see Heit (1998) and Tenenbaum and Griffiths (2001) for some psychological applications of the eliminative approach.

The Descriptive View

In addition to the normative perspective on the diversity principle, there has been a sustained effort by psychologists to document how well the diversity principle serves as a descriptive account of how people carry out informal, inductive reasoning. Osherson, Smith, Wilkie, López, and Shafir (1990) documented diversity effects in adults by using written arguments such as the following:

(1) Hippopotamuses require Vitamin K for the liver to function.
Rhinoceroses require Vitamin K for the liver to function.
___
All mammals require Vitamin K for the liver to function.

(2) Hippopotamuses require Vitamin K for the liver to function.
Hamsters require Vitamin K for the liver to function.
___
All mammals require Vitamin K for the liver to function.

The subjects judged arguments such as (2) to be stronger than arguments such as (1), showing sensitivity to the greater diversity of hippopotamuses and hamsters compared with hippopotamuses and rhinoceroses.

López (1995) devised a stricter test of diversity-based reasoning, in which people choose premise categories rather than simply judge inductive arguments. Subjects were given a fact about one mammal category, and they were asked to evaluate whether all mammals have this property. Subjects were allowed to investigate one other category of mammals. For example, subjects were told that lions have a certain property, then the subjects were asked whether they would test leopards or goats as well. The result was that subjects consistently preferred to test the more dissimilar item (e.g., goats rather than leopards). Hence, people are sensitive to diversity not only when evaluating evidence but also when seeking evidence.

Indeed, a great deal of evidence suggests that adults, mainly Western university students, follow the diversity principle when evaluating written arguments (for a review, see Heit, 2000; for more recent evidence, see Kim & Keil, 2003, and Kincannon & Spellman, 2003; for some well-documented exceptions, see Osherson et al., 1990, and Sloman, 1993). This chapter does not address the controversial question of whether children follow the diversity principle. There is evidence both of children following the diversity principle (Heit & Hahn, 2001; Lo, Sides, Rozelle, & Osherson, 2002) and children not following the diversity principle (Carey, 1985; Gutheil & Gelman, 1997; Lo et al., 2002; López, Gelman, Gutheil, & Smith, 1992).

Challenges to the Normative View

Despite the intuitive appeal of the notion that diverse evidence is strong evidence, unassailable proof of this claim has been elusive. Wayne (1995) made detailed criticisms of both the correlation approach and the eliminative approach. Wayne suggested that the eliminative approach has problems of circularity, claiming that its assumptions regarding elimination of hypotheses are just a redescription of the diversity phenomenon. With reference to the correlation approach, Wayne noted the difficulty of objectively stating whether two sources of evidence are similar or diverse. For example, after the acceptance of Maxwell's electromagnetic theory of light, phenomena that had previously seemed diverse (i.e., magnetic phenomena and optical phenomena) now seemed much more similar. Using Earman's (1992) own derivations of the diversity principle, Wayne also showed that exceptions to the diversity principle can exist—namely, that nondiverse observations can lead to strong inferences if this evidence is nonetheless very surprising. That is, "an unexpected pair of results which are highly correlated can boost the probability of an hypothesis more than a pair of diverse results with relatively high priors" (p. 114). By way of example, Wayne pointed to the near-simultaneous discovery in 1974 of a previously unknown subatomic particle in two laboratories; this, he argued, was a case of nondiverse evidence with strong implications for a revision of theories in physics.

Indeed, Lo et al. (2002) raised a closely related criticism of the normative status of the diversity principle. They too argued that what is crucial is not the diversity of observations but rather the surprisingness of observations. Lo and colleagues referred to this principle for evaluating evidence as the *premise probability principle.* They also suggested a set of exceptions to the diversity principle, such as the following:

(3) Squirrels can scratch through Bortex fabric in less than 10 seconds.
Bears can scratch through Bortex fabric in less than 10 seconds.

All forest mammals can scratch through Bortex fabric in less than 10 seconds.

(4) Squirrels can scratch through Bortex fabric in less than 10 seconds.
Mice can scratch through Bortex fabric in less than 10 seconds.

All forest mammals can scratch through Bortex fabric in less than 10 seconds.

It seems intuitive that squirrels and bears are a more diverse pair than squirrels and mice. Yet Lo et al. argued that (4) is stronger than (3) because the evidence about squirrels and mice is more surprising than the evidence about squirrels and bears. That is, the knowledge that small animals are less capable of feats of strength than are large animals makes the evidence about squirrels and mice more surprising than the evidence about squirrels and bears.

Our own reaction to these exceptions to the diversity principle suggested by Wayne (1995) and Lo et al. (2002) is that they are indeed exceptions but they by no means undermine the normative status of the diversity principle itself. In the example of the discovery of a new subatomic particle in 1974, physicists were influenced not only by diversity but also by many other sources of knowledge in particle physics. In the example of scratching through Bortex fabric, people would be influenced not only by diversity but also by other knowledge about animals and their strength. In other words, these exceptions as stated do not contain all the premises on which the arguments are based. Reasoning about these arguments is also influenced by other hidden premises or background knowledge so that diversity is not being assessed in isolation. Therefore these counterexamples do not invalidate the diversity principle because they are not pure tests of diversity. Rather, they show that people will use other knowledge when possible.

By the very nature of inductive reasoning, it is always normative to consider other knowledge (Skyrms, 2000). Inductive inferences are never 100% certain; hence, it is always possible to improve inductive inferences by the application of further knowledge. Indeed, philosophers of science have not claimed that the diversity principle is the sole principle for assessing evidence. For example, Popper (1963, p. 232) listed diversity of supporting evidence as one of six criteria for assessing a scientific theory and followed this with a discussion (p. 240) of the importance of considering other sources of background knowledge when testing a theory.

In sum, the exceptions to the diversity principle suggested by Wayne (1995) and Lo et al. (2002) are valuable because they illustrate that notwithstanding the normative status of the diversity principle, it is also normative to consider other sources of knowledge when making an inductive inference. With this point made, we now return to the descriptive view of the diversity principle, which also underscores the importance of considering other sources of knowledge.

Challenges to the Descriptive View

The evidence for people following the diversity principle has mainly accrued from experiments on Western college students, and indeed there is a great deal of evidence from such sources. However, when other subject populations are considered and evidence is collected at a greater distance from the psychology lab, exceptions to the diversity principle emerge. In their study of Itza'–Mayan adults from the rainforests of Guatemala, López, Atran, Coley, Medin, and Smith (1997) did not find evidence for diversity-based reasoning when they used arguments with various living things and questions about disease transmission. Indeed, sometimes the Itza' reliably chose arguments with nondiverse premise categories over arguments with diverse categories. Apparently, they were using other knowledge about disease transmission that conflicted with diversity-based reasoning. For example, given a nondiverse argument—that two similar kinds of tall palm trees get a certain disease—one person claimed that it would be easy for shorter trees, located below the trees in question, to get the disease as well.

Giving further support to this idea that other strategies and knowledge can overrule diversity, Proffitt, Coley, and Medin (2000) reported that American adults who are tree experts (such as landscapers and park maintenance workers) did not show strong diversity effects when reasoning about trees and their diseases. The tree experts seemed to be relying on the knowledge that tree diseases tend to spread readily within tree families such as elms and maples. Their inferences seemed to follow an alternate strategy that did not assess diversity against the broad category of "all trees" but rather considered the size of various tree families.

Again, our reaction to these exceptions to the diversity principle is that they do not actually invalidate the diversity principle but rather show the use of other knowledge. It is plausible that the landscapers discussed by Proffitt et al. (2000) would still show diversity effects for other stimuli. Indeed, in a follow-up study, López et al. (1997) found that the Itza' do show diversity effects on other questions. For example, they were told to imagine buying several bags of corn, in a similar problem to that of Nagel (1939). The question was whether it would be better to inspect two corn cobs from one bag or one corn cob from each of two different bags, and, indeed, the Itza' showed a diversity effect.

In a recent paper, Medin, Coley, Storms, and Hayes (2003) documented further exceptions to the diversity principle. Some of these exceptions involve diversity being overridden by other knowledge, particularly causal knowledge. However, one phenomenon, referred to as *nondiversity by property reinforcement*, potentially makes a more direct challenge to the diversity principle that is not as easily explained in terms of the use of other knowledge. The idea behind nondiversity by property reinforcement is that two diverse categories may nonetheless have some characteristic in common and tend to generalize only to other categories with this same characteristic. According to the nondiversity by property reinforcement effect, "if an otherwise diverse set of premises shares a salient property not shared by the conclusion category, the reinforcement of the property might weaken that argument relative to a related argument with less diverse premises" (p. 523). This phenomenon is illustrated by the following example:

(5) Polar bears have property X.
Antelopes have property X.

All animals have property X.

(6) Polar bears have property X.
Penguins have property X.

All animals have property X.

When given a forced choice between polar bears and antelopes versus polar bears and penguins, subjects judged the two animals from the same biological class, polar bears and antelopes, to be more similar than the two animals from different biological classes, polar bears and penguins. However, when asked to assess the inductive strength of each argument, argument (6) was judged to be less convincing than argument (5)—that is, argument (5) had less diverse evidence, yet it was the stronger argument. It seems intuitive that although polar bears and penguins are from different biological classes, they still share the characteristic of living in a cold climate. Property *X* does not seem to extend to all animals but applies only to animals that share the characteristic of living in cold climates.

Medin, Coley, Storms, & Hays (2003) investigated this nondiversity by property reinforcement effect using several stimulus sets and did find significant evidence overall for this phenomenon. However, the results were not always consistent. Sometimes the similarity comparisons yielded results that were the opposite of those anticipated—that is, sometimes same-class animals were judged to be more diverse than different-class animals. Sometimes the inductive-strength judgments did go in the direction of diversity rather than nondiversity. Hence, in recent experiments, we (Heit & Feeney, in press) have conducted further tests of the nondiversity by property reinforcement phenomenon, using Medin and colleagues' stimuli as well as other materials. In general, we followed the procedure used by Medin et al. except we collected similarity judgments in a different way. Rather than asking subjects to make a forced choice between two same-class animals and two different-class animals, we asked subjects to make individual similarity ratings corresponding to each of the arguments. This procedure facilitated the key analysis, which examined the correlation between similarity and inductive strength, allowing consideration of the whole pattern of results.

Experiment 1

The first experiment had two groups of subjects: 72 subjects made judgments of inductive strength, and 45 subjects made similarity ratings.

The experiment used seven pairs of inductive arguments, adapted from items in Medin et al. (2003), as shown in Table 6.1. Each pair included an argument based on two animals from the same biological class (such as penguin–eagle, both birds) and an argument based on two animals that belonged to different biological classes (such as penguin–polar bear, one bird and one mammal) or were distantly related within a biological class. The different-class animals nonetheless shared certain salient characteristics (such as living in a cold habitat for penguin–polar bear). The first five pairs in Table 6.1 had been validated in terms of similarity judgments collected by Medin et al.—that is, when given a forced choice between whether the same-class animals were more similar or the different-class animals were more similar, subjects tended to say that the same-class items were more similar. For example, the majority of people stated that penguin–eagle had more similarity than penguin–polar bear. The final two pairs showed the opposite pattern of similarity judgments.

Table 6.1. Stimuli and Results for Experiment 1

Same class	Inductive strength (%)	Similarity	Different class	Inductive strength (%)	Similarity
Penguin–eagle	43.6	3.18	Penguin–polar bear	40.1	4.18
Kangaroo–elephant	50.1	2.16	Kangaroo–frog	49.4	3.02
Camel–rhino	52.9	2.98	Camel–desert rat	39.5	4.18
Polar bear–antelope	52.6	2.16	Polar bear–penguin	40.1	4.18
Chimpanzee–cow	51.2	1.89	Chimpanzee–dolphin	55.9	2.27
Bat–elephant	51.1	1.69	Bat–robin	40.0	4.07
Pig–whale	61.3	1.64	Pig–chicken	48.3	3.53
MEAN	51.8	2.24	MEAN	44.7	3.63

The inductive arguments were given as part of a pen-and-paper survey. The questions were of the following form:

Given the facts that:

Penguins have property *X*.
Antelopes have property *X*.

How likely is it that:

All animals have property *X*?

Note that whereas Medin et al. (2003) had actually used a variety of conclusion categories, we consistently used animals so as to facilitate the correlational analyses across items. Although the polar bear–penguin argument was essentially used in two different pairs by Medin et al., we collected data for this argument only once. Subjects were asked to respond to each question on a 0%-to-100% scale. Half the subjects were also asked to justify each judgment on an additional line, but because this information did not affect the results, we do not report on this further.

The similarity condition used a pen-and-paper survey containing the same 13 pairs shown in Table 6.1. Subjects were asked to make similarity ratings on 1-to-9 scale, with higher numbers indicating greater similarity.

We now turn to the results. The inductive strength ratings for each argument are shown in Table 6.1. Overall, there were significantly higher ratings for same-class arguments (51.8%) than for different-class arguments (44.7%). The results were fairly consistent across items. That is, the same-class argument had greater strength than the different-class argument for six of seven pairs. Note that Medin et al. (2003) also found higher inductive strength ratings for same-class arguments.

The similarity ratings for each argument are also shown in Table 6.1. Overall, there were significantly higher similarity ratings for the different-class arguments (3.63) than for the same-class arguments (2.24). Indeed, each pair of arguments showed this same pattern. Note that here our results depart from those of Medin et al. (2003), who found in the first five pairs a tendency to say that the same-class arguments were more similar than the different-class arguments. When the inductive-strength ratings and the similarity ratings are combined and same-class arguments are compared with different-class arguments, it appears that there was a diversity effect overall, rather than a nondiversity effect—that is, the same-class arguments were judged as more diverse, in terms of having lower similarity ratings, and were also judged as being inductively stronger.

Finally, we examined the correlation between inductive ratings and similarity judgments, taken over the 13 unique arguments. There was a statistically significant negative correlation, $r = -0.86$. That is, when the two animals in an inductive argument were judged as more diverse, in terms of having lower similarity ratings, the inductive strength of that argument tended to be higher. Hence, the correlational analysis also showed a diversity effect.

In sum, this experiment replicated the results obtained by Medin et al. (2003) in terms of the inductive judgments but revealed a different pattern for the similarity judgments; therefore the overall interpretation was a diversity effect rather than a nondiversity effect. The differences in similarity judgments, at a theoretical level, are a good illustration of the dynamic and context-dependent nature of similarity (Medin, Goldstone, & Gentner, 1993). From a methodological standpoint, they could reflect the effects of asking for similarity judgments in different ways (for further discussion, see Heit & Feeney, in press.) In light of the different similarity results for this stimulus set, we thought it would be valuable to investigate the nondiversity by property reinforcement effect with another stimulus set.

Experiment 2

The link between Experiment 1 and Experiment 2 is that they both used stimuli for which nondiversity by property reinforcement would be predicted. Experiment 2 used a stimulus set adapted from Heit and Rubinstein (1994), who had also created pairs of arguments contrasting animals from the same biological class to animals from different biological classes (see Table 6.2). For example, bears and whales are both mammals. These were contrasted with tunas and whales, which are in different biological classes but share the characteristic of living in the sea. Heit and Rubinstein collected two kinds of similarity ratings for these stimuli: similarity with respect to anatomy and similarity with respect to behavior. In effect, asking for similarity judgments with respect to anatomy would encourage subjects to ignore other associations. As shown in Table 6.2, the same-class animals were overall generally considered more similar in terms of anatomy than the different-class animals. Hence, these stimuli are compatible with Medin and colleagues' (2003) own aims of looking for nondiversity effects when same-class animals are considered more

Table 6.2. Stimuli and Results for Experiment 2

Same class	Inductive strength (%)	Similarity	Different class	Inductive strength (%)	Similarity
Bear–whale	59.8	3.29	Tuna–whale	22.8	5.56
Mouse–bat	35.1	4.99	Sparrow–bat	31.8	4.81
Lizard–snake	21.7	6.47	Worm–snake	20.3	4.90
Trout–shark	19.0	6.56	Wolf–shark	59.7	2.32
Robin–hawk	16.2	7.29	Tiger–hawk	64.8	2.29
Grasshopper–mosquito	21.0	4.43	Vampire bat–mosquito	34.8	3.19
Ant–bee	24.2	5.15	Hummingbird–bee	35.3	3.40
MEAN	28.1	5.45	MEAN	38.5	3.78

similar than different-class animals. However, Heit and Rubinstein found that the different-class animals were considered somewhat more similar in terms of behavior than were the same-class animals. Hence, in these stimuli the different-class animals nonetheless share some characteristics.

The inductive arguments in Experiment 2 had the following form:

> Given the facts that:
>
> Bears have property *X*.
> Whales have property *X*.
>
> How likely is it that:
>
> All animals have property *X*?

This experiment was conducted with 30 subjects. The inductive-strength ratings for each argument are shown in Table 6.2. Overall, there were significantly higher ratings for different-class arguments (38.5%) than for same-class arguments (28.1%). However, the results were somewhat inconsistent across items—that is, the different-class argument had greater strength than the same-class argument for just four of seven pairs. Still, there was an overall diversity effect with respect to anatomical similarity: The different-class arguments had been judged as more diverse, in terms of having lower anatomical similarity ratings, in Heit and Rubinstein (1994); they were judged here as being inductively stronger.

Next, we examined the correlation between inductive ratings and similarity judgments, taken over the 14 separate arguments. The correlation between inductive strength and anatomical similarity was –0.85. That is, when the two animals in an inductive argument were judged as more diverse, in terms of having lower anatomical similarity ratings, the inductive strength of that argument tended to be higher. Hence, the correlational analysis also showed a diversity effect in terms of anatomical similarity. Note that the correlation value is almost the same as in Experiment 1, despite the fact that the similarity judgments were obtained at the University of Michigan, Ann Arbor, U.S.A., in 1992, and the inductive strength judgments were collected at the University of Warwick, U.K., in 2003. Furthermore, the correlation between inductive strength and behavioral similarity was 0.07, not significantly different from zero.

Heit and Rubinstein (1994) showed that the property itself in an inductive argument plays a crucial role (see also Ross & Murphy, 1999). Namely, if the property being inferred concerned animals' behavior rather than anatomy, subjects tended to assess inductive strength not only in terms of anatomical similarity but also in terms of behavioral similarity. The properties used in Medin et al. (2003) included internal characteristics such as "contains retinum" and scientific-sounding designations such as "has property *X12*" and plausibly were interpreted as having to do more with anatomy than behavior. Likewise, in our own Experiments 1 and 2, it is plausible that Property C and so on were interpreted as anatomical rather than behavioral. Therefore in a third experiment we examined the role of diversity when subjects were making judgments about behavioral properties. This experiment was almost identical in method to Experiment 2 except that the property being inferred was emphasized to be behavioral. On every stimulus item, the term "behavioral property" was used, as in "Behavioral Property C." In addition, the instructions stated that "properties refer to behavioral characteristics such as movement, eating habits, and food-gathering or hunting techniques." In general, this third experiment yielded results similar to those obtained in Experiment 2. Taken across the 14 stimulus items, the correlation between inductive-strength judgments in the third experiment and in Experiment 2 was 0.92. Likewise, in the third experiment, there was a negative correlation between inductive strength and anatomical similarity, –0.64, and a near-zero correlation between inductive strength and behavioral similarity, –0.07. We interpret this third experiment as showing the robust nature of the diversity effect with respect to anatomical similarity, even when the properties inferred concern behavior.

Hence, in three new experiments, we have found diversity effects rather than nondiversity effects. There were substantial theoretical reasons to predict a nondiversity by property reinforcement effect, based on Medin et al.'s (2003) relevance theory of induction, which provides an explanation in terms of people looking for distinctive properties of premise categories on the assumption that these categories were not chosen randomly but instead were presented as part of a discourse. Yet the diversity effect appears to be robust overall. Still, we would not rule out the possibility of some nondiversity items being found against a larger pattern of diversity effects overall. Medin et al. did document several other systematic exceptions to diversity and related effects. For example, strong evidence supported the use of causal knowledge to override the diversity effect in cases that had been predicted by the relevance theory proposed by Medin et al. Overall, a great deal of evidence supported the relevance theory in other results reported by Medin et al.

Conclusion

At a general level, work on the diversity principle shows how a seemingly straightforward idea, that diverse evidence will be strong evidence, turns into a rich area of research when considered from multiple perspectives taking in the history and philosophy of science, statistical theory, experimental psychology, developmental psychology, and cross-cultural psychology. Although this chapter has been presented as a defense of the diversity principle, our ultimate aim is not to find a monolithic, yes-or-no answer to whether the diversity principle succeeds as a normative and a descriptive account. Both the successes and the failures of the diversity principle have proved to be theoretically revealing about the nature of inductive reasoning and its relations to other important topics such as categorization, similarity, and the influences of background knowledge on cognition.

References

Bacon, F. (1898). *Novum organum*. London: George Bell & Sons. (Original work published 1620)

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64,* 159–174.

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, England: Oxford University Press.

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

Heit, E., & Feeney, A. (in press). Relations between premise similarity and inductive strength. *Psychonomic Bulletin and Review.*

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43,* 243–273.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 411–422.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice Hall.

Horwich, P. (1982). *Probability and evidence.* Cambridge, England: Cambridge University Press.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach.* Chicago: Open Court.

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

Kincannon, A., & Spellman, B. A. (2003). The use of category and similarity information in limiting hypotheses. *Memory and Cognition, 31,* 114–132.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 16,* 181–206.

López, A. (1995). The diversity principle in the testing of arguments. *Memory and Cognition, 23,* 374–382.

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

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63,* 1070–1090.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin and Review, 10,* 517–532.

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

Nagel, E. (1939). *Principles of the theory of probability.* Chicago: University of Chicago Press.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge.* London: Routledge.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 811–828.

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

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Skyrms, B. (2000). *Choice and chance: An introduction to inductive logic* (4th ed.). Belmont, CA: Wadsworth.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–641.

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science, 62,* 111–121.