

From symptoms to causes: Diversity effects in diagnostic reasoning

NANCY S. KIM and FRANK C. KEIL
Yale University, New Haven, Connecticut

A single causal agent can often give rise to a cascade of consequences that can be envisioned as a branching pathway in which symptoms are the terminal nodes. In three studies, we investigated whether reasoning about root causes on the basis of such symptoms would conform to a diversity effect analogous to that found in inductive reasoning about properties of hierarchically organized categories. A strong diversity effect was found both for reasoning about medical diseases that drew on existing background knowledge and for reasoning that did not. Specifically, the presence of a root cause was more likely to be induced when the symptoms present were further apart in the branching structure.

In many aspects of life, we are confronted with diagnostic questions. Diagnoses most obviously occur in medical settings, where patients present with a set of symptoms and a clinician makes an induction about the likely disease, or original cause. A car mechanic goes through a similar procedure, trying to determine which primary cause is responsible for such symptoms as a coughing engine, poor acceleration, or an unusually smoky exhaust. Diagnoses also occur somewhat more subtly in many other everyday settings. Whether figuring out why a computer is not functioning properly, why an undergraduate is flunking out of college, or why a company is going bankrupt, we are frequently confronted with situations in which a wide range of “symptoms” is presented and our task is to figure out a likely cause or causes. The same problem occurs for positive as well as negative sets of events, such as puzzling out why one undergraduate did so well, why a company was so profitable, or why a person is resistant to a disease that all those around him or her have acquired.

In many of these cases, it is common for people to look for a single primary cause that gave rise to a cascade of consequences. This manner of drawing inferences raises the question of what sorts of information might lead people to assume a given single cause was particularly likely, given a constellation of subsequent events. Many factors can influence judgments of the presence and strength of causal agents. These factors may range from the known prior and conditional probabilities of features (Jensen, 1996; Rehder & Hastie, 2001) to the plausibility of a given cause–effect relationship (Ahn & Kalish, 2000). There are, however, structural properties of causal relations that

may influence such judgments above and beyond effects that are due to the mere content of the events involved. For example, the causal status effect shows that cause features are deemed more central to a category than their effects, even if their frequencies of occurrence are identical (Ahn, Kim, Lassaline, & Dennis, 2000). Such findings have launched important discussions on how the perception of causality might influence reasoning and categorization.

In this paper, we explore the effects of a different aspect of causal knowledge: the influence of causal hierarchies on diagnostic reasoning. Our use of the term *causal hierarchies* in this paper refers to cases in which a single cause is perceived as having a branching set of consequences. (In some causal systems, the intermediate consequences can converge to cause a common effect. In contrast, the present paper explores cases in which the branching set of consequences continues to split without rejoining.) To the extent that many of our beliefs about complex systems may be represented by causal hierarchies, it is important to understand whether there are ways in which such beliefs influence reasoning. One powerful influence may arise from a *diversity effect*—namely, that knowing information about a diverse set of locations as defined by the causal hierarchy allows for broader inductive inferences than does knowing information about a *proximal* set of locations.

Later in the paper, we will argue that diversity effects are likely to extend beyond diagnosis to many other cases of inductive reasoning, but diagnosis situations pose one of the most direct tests of the effect. Furthermore, they offer the closest analogy to well-known diversity effects arising from the hierarchical structuring of categories, which we will describe in the next section. Drawing on such an analogy, it is possible to make predictions about how a hierarchically structured set of causal beliefs about a disease might influence diagnoses.

Category-Based Diversity Effects

As an example of category-based diversity effects, consider the tree shown in Figure 1, which shows the hi-

This research was supported in part by a National Science Foundation Graduate Research Fellowship to N.S.K. and National Institute of Health Grant R01-HD23922 to F.C.K. We thank Lee Brooks, James Hampton, Evan Heit, Geoffrey Norman, and Barbara Spellman for helpful comments concerning this project. Correspondence concerning this article should be addressed to N. S. Kim or F. C. Keil, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06520-8205 (e-mail: nskim@aya.yale.edu or frank.keil@yale.edu).

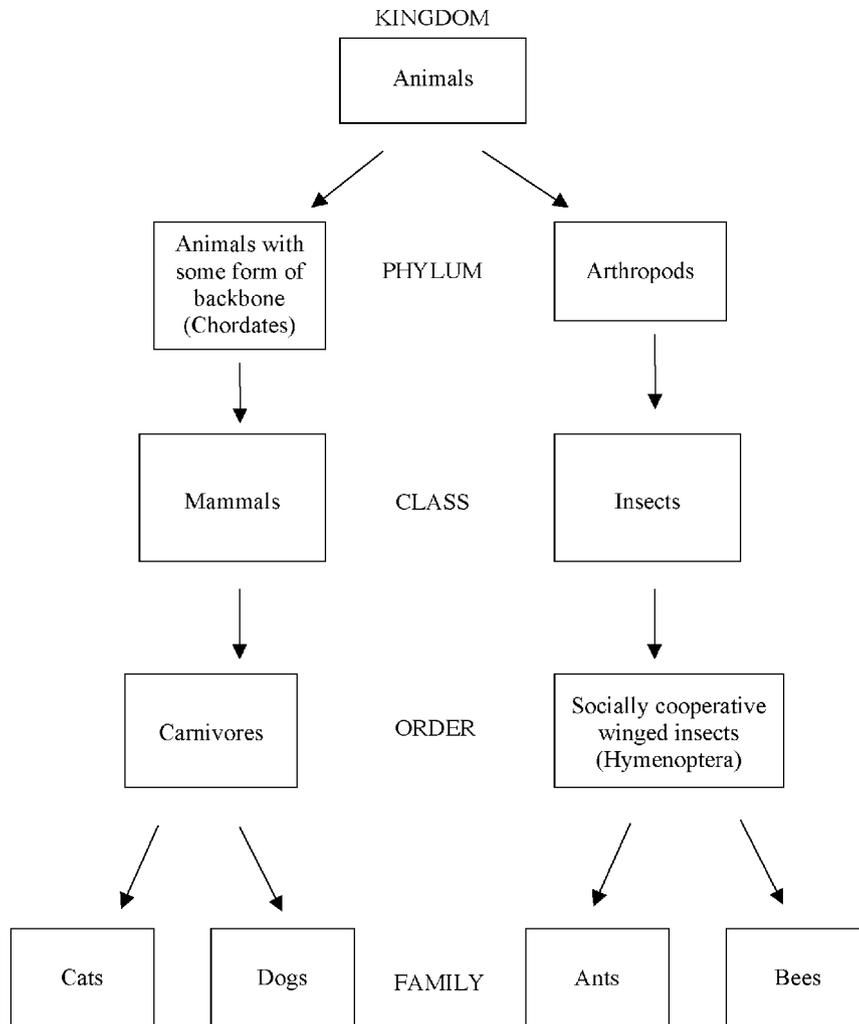


Figure 1. Hierarchically arranged categories depicting a portion of the animal kingdom.

erarchical organization of some animal categories. If participants are presented with a property that is true of two animals that are far apart on the tree and, therefore, more diverse, they will judge the property as more likely to be true of all animals than if it is true of adjacent members of a category. For example, if told that dogs and ants both have homeobox genes, one is much more likely to assume that all animals do than if one is told that bees and ants do. People seem to assume that the more distinct two categories are from each other, the more general any property is that is true of both of them (Osherson, Smith, Wilkie, López, & Shafir, 1990). This pattern of reasoning seems to rest on the notion that members of the most superordinate category tend to have a major characteristic in common (i.e., all animals are living kinds).

The initial research carried out on category-based diversity effects was developmental (Carey, 1985). In that work, which tested participants' willingness to induce

invisible properties about living kinds, the diversity effect seemed to be much weaker in 6-year-old children than in adults. Carey argued that this finding indicated that younger children have a less coherent view of the category of living things. This insufficient-biological-theory explanation, however, has been controversial given that children demonstrate a more coherent sense of biology on a range of nondiversity induction tasks (Gutheil, Vera, & Keil, 1998; Inagaki & Hatano, 2002). Although Carey argued that performances on the diversity induction tasks and nondiversity induction tasks were both based on the extent to which children had a coherent underlying theory of biology, the tasks may in fact draw on somewhat different cognitive resources. Indeed, when subsequent studies (e.g., López, Gelman, Gutheil, & Smith, 1992) also showed weakened category-based diversity effects in children, a developmental shift was proposed. Proponents of this developmental shift

hypothesis suggested that children shifted from initially being unable to use diversity in inductive judgments to acquiring that skill as late as the 2nd decade of life (Gutheil & Gelman, 1997).

The notion that diversity-based reasoning might be a sophisticated and relatively late developing cognitive ability, however, no longer seems plausible given more recent findings. For instance, if the properties being induced are salient relational ones, such as who would be more likely to play with a novel object, children as young as 5 years of age show robust diversity effects (Heit & Hahn, 2001). These findings suggest that for categories, diversity-based reasoning may in fact emerge very early on, possibly as early as preschool, given the appropriate tasks.

In contrast, it is well established in adults that there is a diversity effect for reasoning about properties of cate-

gories that are hierarchically arranged. This effect has been shown extensively in studies on adult populations (e.g., López, 1995; Osherson et al., 1990). The question for the present paper is whether these findings may also be transferred to an analogous type of knowledge structure.

Causally Based Diversity Effects

Consider the analogous structure for a disease causation network shown in Figure 2. Figure 2 represents the knowledge that an individual might have of how radiation sickness is caused. In many cases, as will be argued below, that individual may not even be aware that the knowledge forms such a hierarchy, having only considered at different times particular linear pathways from the root cause to each final symptom. Nonetheless, over time, the hierarchy may be formed implicitly and, subsequently, may in-

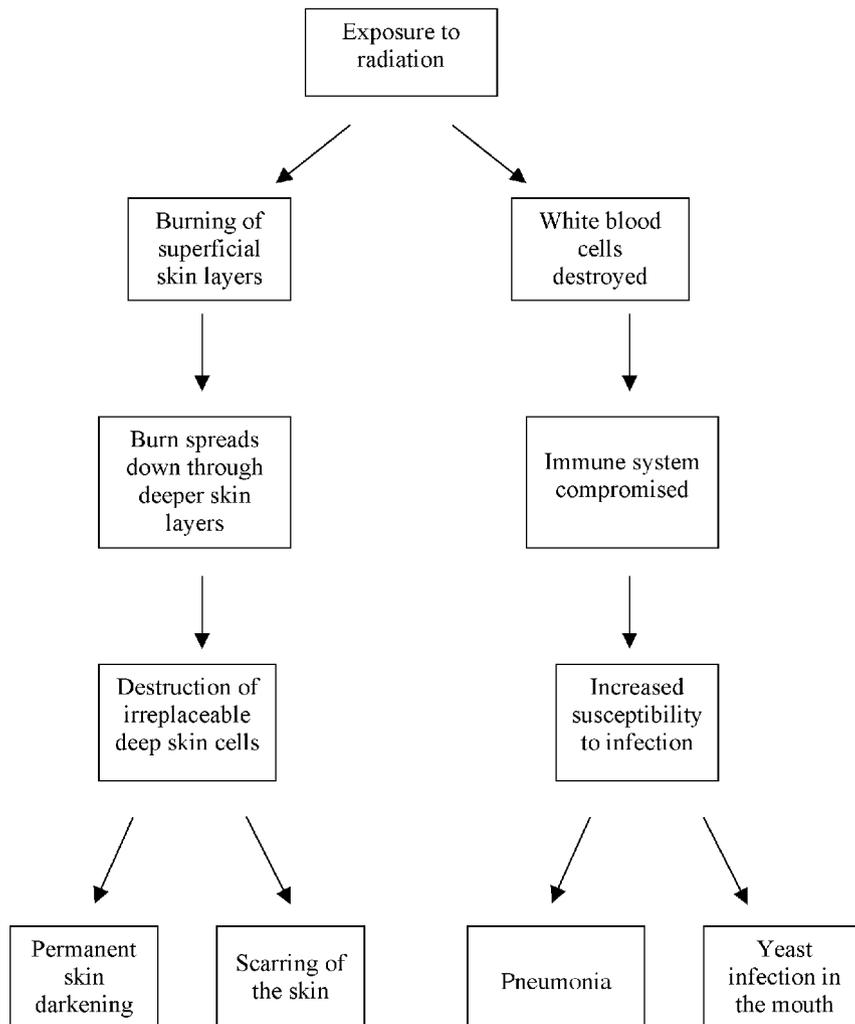


Figure 2. Summary causal diagram of the radiation sickness stimuli used in Study 1. For this medical condition, the participants were asked to judge whether a patient with pneumonia and scarring of the skin or a patient with pneumonia and a yeast infection in the mouth was more likely to have radiation sickness.

fluence reasoning. Consider two cases as an example. In the first case, a patient is described as having the symptoms of a yeast infection in the mouth and pneumonia. In the second, a patient is described as having permanent skin darkening and a yeast infection of the mouth. A person who had the hierarchical causal knowledge represented in Figure 2 might not have much information about the base rates of these symptoms, perhaps having only heard about one instance of each. Nonetheless, the first case might seem a less likely instance of the radiation poisoning than the second. The causal diversity hypothesis would state that an original cause is likely to be implicated to the extent that observed consequences arise from a higher common node in the hierarchical causal tree. Thus, two symptoms whose first common node is directly above them (the yeast infection, pneumonia case) would be less likely to be seen as the result of a higher original cause than would two symptoms whose first common node is higher in the tree (as in the yeast infection, permanent skin darkening case). This difference would occur because the simplest, most parsimonious explanation for the two proximal symptoms would be a lower node in the tree, whereas given the hierarchical causal tree, a simple, single coherent explanation for the two diverse symptoms can be found only at the highest point of the tree (Read & Marcus-Newhall, 1993; Thagard, 1992).

If the causal diversity hypothesis is true, it may be that consequences with lower common nodes are more easily construed as arising from different original causes. For instance, yeast infection and pneumonia can arise from a chain of three possible causes that precede it in a linear chain: white blood cells being destroyed, compromising of the immune system, and increased vulnerability to infections. Other, still higher causes could give rise to each of these, elevating the possibility that these two symptoms together do not necessarily indicate a diagnosis of radiation poisoning. By contrast, yeast infection and permanent skin darkening have radiation exposure as their only common cause. Of course, other factors could independently cause each of these symptoms, but it seems less plausible that they could occur simultaneously in cases not sharing a higher common cause.

Thus, in a manner analogous to the influence of hierarchically nested categories on property induction, hierarchically organized causal intuitions are expected to give rise to a diversity effect in diagnostic reasoning. The analogy between the causal diversity hypothesis and the diversity effect found for category hierarchies, like all analogies, suggests a common relation in the face of salient differences; and indeed, the differences between inductions about hierarchically embedded categories and about a hierarchy of branching causal processes are substantial. With categories, one is making inductions about the likelihood of a property's being true of a category. In diagnostic reasoning, in contrast, one is making inductions about the likely original cause of final symptoms or effects. Thus, one form of reasoning is about entities and

their properties, whereas the other is about events that unfold over time.

However, there is good reason to suspect that diversity effects will occur for causal hierarchies, as well as for hierarchies of categories. Both types of diversity-based effects may be traced back to scientific reasoning's emphasis on converging evidence. For example, in arguing for the possibility that younger children are capable of diversity-based reasoning, Heit and Hahn (2001) pointed to the long-recognized value of diverse observations in the sciences (see also Spellman, López, & Smith, 1999). At least since the time of Bacon, philosophers of science have noted that more diverse observations are more valuable in supporting hypotheses (Heit & Hahn, 2001). Many branches of the sciences stress the importance of converging forms of evidence, an emphasis that is often an indirect argument for the diversity effect (i.e., support for a theory is perceived as most powerful if corroborating evidence is gleaned using very different tasks and measures). In other words, the wider the angle of convergence, the stronger the evidence. This notion was eloquently argued by the 19th century philosopher of science and polymath William Whewell, who described a "consilience of inductions" in which major advances in science, such as the Copernican revolution, were fueled by support coming from highly diverse sources (Whewell, 1840/1999, 1847). Whewell's proposal has been embraced on many occasions in discussions of the extraordinarily diverse support for evolutionary theory (e.g., Magnus, 1996; Ruse, 1998).

In short, the widely held notion that the value of evidence is great if its sources are dissimilar leads us to expect diversity effects in both categorization and causal reasoning. In the Discussion section, we will consider in more detail some potential psychological mechanisms responsible for diversity effects and the extent to which they might be shared across these two forms of reasoning (i.e., category-based and causally based diversity effects).

STUDY 1

Real-Life Medical Conditions

Method

Participants. Forty Yale University undergraduate students participated in this study in exchange for either partial fulfillment of the requirements for an introductory psychology course or monetary compensation at the rate of \$7 per hour. Twenty of the students completed the experimental task, and 20 completed a control task.

Materials. Six sets of medical stimuli were developed for this study. The following medical conditions were selected because they include a variety of symptoms arising from distinct causal pathways: allergic reaction, ingestion of a specific fungus, pregnancy, radiation sickness, Type II diabetes, and rheumatic fever. The allergic reaction and fungal ingestion materials were based on real-life medical conditions, with minor changes being made to fit them into hierarchical structures of the type shown in Figure 2. The remaining four medical conditions fit naturally into that hierarchical structure and did not require any adjustment. For each medical condition, a summary causal diagram was drawn up. Each summary

diagram depicted the disease as stemming from a single cause that immediately branched out to launch two separate causal chains. Each of these two causal chains contained three components, the last of which gave rise to two distinct symptoms. Thus, each disease resulted in four symptoms in all (see Figure 2 for a sample diagram).

The summary diagrams themselves were not utilized in the experimental task, because they inherently contained spatial distance cues that could conceivably affect participants' judgments. Therefore, for the participants' task, a set of four separate diagrams was created from each summary diagram. Each of these four diagrams depicted the deepest-cause symptom as causing the next three components of a chain and one of the four symptoms. Thus, each set of four diagrams for each disease contained the same information as the single summary diagram for that disease, while eliminating all spatial distance cues. Each diagram was labeled at the top with the name of the medical condition, and each set of four diagrams was also numbered consecutively.

For each medical condition, a sheet containing the dependent measure question was also constructed. The top of the sheet was labeled with the name of the condition, and two statements followed concerning the symptoms of two medical patients. Each medical patient was said to have two symptoms. One symptom was identical in the two patients. However, in one of the patients, the second symptom was a product of the same causal chain as the first symptom, and in the other patient, the second symptom was a product of the other causal chain. Below these two statements, there appeared the question, "Who, in your opinion, is more likely to have [Condition X]?"

Design and Procedure. The participants completing the experimental task were given a booklet containing instructions, the causal diagram sets, and the dependent measures. On the first page, the participants were informed that they would read information in the booklet concerning six different medical conditions. They were told that for each medical condition, they would read four different diagrams, each illustrating how a specific symptom of that condition comes about. They were asked to read the information carefully. Then, they were told that after reading the four diagrams for a condition, they would be asked to integrate all that information into a single summary paragraph and then to answer an application question about the medical condition. They were told that they would go through this process for all six medical conditions and that they could address their questions to the experimenter.

For each medical condition, the participants first viewed the four diagrams, each on a separate page. They then turned to a page that instructed them to write a single paragraph explaining all the information about that medical condition that they had read in the preceding diagrams. They were told that they were allowed to refer back to the diagrams as needed. Then they were presented with the sheet containing the dependent measure question for that medical condition. Finally, they turned to a sheet informing them that they had finished the task for this medical condition and to turn the page to begin reading the information about the next medical condition.

For each participant, the diagrams were arranged in one of two pseudorandom orders. These orderings were determined randomly, except that diagrams containing symptoms stemming from the same causal chain were not placed consecutively. The order in which the two patients were presented on the dependent measure sheets was also counterbalanced between subjects. Finally, the order in which the six medical conditions were presented was randomized for each participant.

The participants completing the control task went through an identical procedure, except that they did not see the causal diagrams and did not write a paragraph summarizing the causal information. Instead, for each medical condition, they learned the name of the condition and the four symptoms that they were told were "known to be associated with [Condition X]." Then, for each condition, they completed the dependent measure sheet used in the experimental task.

Results and Discussion

Data were coded such that an assignment of "1" denoted an answer based on the diversity of symptoms as defined by the experimental task causal diagrams given to that particular participant, "0" denoted the other answer, and "0.5" denoted that the participant indicated that he or she was unable to choose. The participants chose the "0.5" option only 7.1% of the time. Six data points were thereby supplied by each participant, one for each of the six medical conditions. A 2 (task: experimental or control) \times 2 (diagram order: 1 or 2) \times 2 (patient order: A or B) analysis of variance (ANOVA) was conducted. The data were collapsed across the six medical conditions to meet distributional requirements for ANOVAs.

There was a significant main effect of task, such that the participants completing the experimental task based their responses significantly more on the diversity of symptoms ($M = .850$, or 85.0%, of the responses, $SD = .187$) than did the participants in the control task [$M = .578$, or 57.8%, of the responses, $SD = .165$; $F(1,32) = 24.5$, $MS_e = 0.03$, $p < .001$]. Broken down by disorder, the results appear to run in the same direction for five of the six medical conditions (see Figure 3). No other main effects or interactions with task neared significance ($p > .6$).

One-sample t tests were also conducted to compare responses from the experimental group with chance levels (50%) and control responses with chance. Experimental participants chose the diversity-based response significantly more than chance [$t(19) = 8.39$, $p < .001$]. In addition, control participants tended to choose the diversity-based response more than chance [$t(19) = 2.06$, $p < .06$], indicating that it is likely that the participants may have had some degree of prior knowledge about the medical conditions. However, this does not undermine the present findings, because the participants given explicit, specific causal knowledge in the experimental condition showed the diversity effect to a degree well surpassing that shown by the participants in the control condition. In addition, as will be seen in Studies 2 and 3, the diversity effect still appears robustly when the problem of prior knowledge is completely controlled.

Overall, Study 1 supports our main hypothesis that people will base their inductive reasoning concerning medical diagnosis on how diverse the symptoms are. Four of the six medical conditions fit naturally into the hierarchical structure depicted in Figure 2, and the remaining two conditions required only minor modifications to fit the structure, suggesting that the effect occurred for diseases with high ecological validity.

STUDY 2 Realistic Artificial Diseases

Study 1 showed a strong diversity-based within-category reasoning effect in real-life medical conditions. However, the disadvantage of using real-life categories was in controlling the materials; it was virtually impossible to

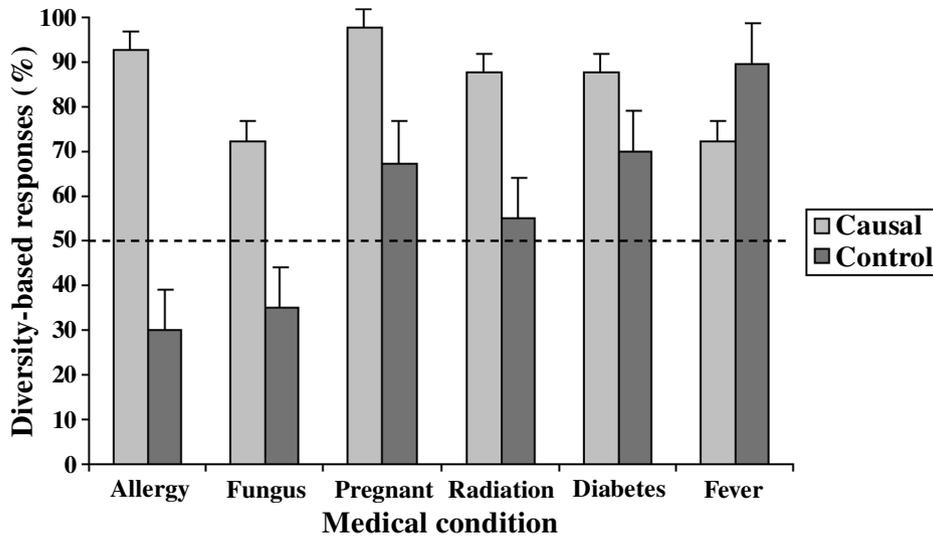


Figure 3. Percentages of diversity-based reasoning in the causal versus control conditions of Study 1, broken down by medical condition. The horizontal dotted line denotes what responses would be at chance levels. Error bars indicate standard errors.

create completely balanced real-life stimuli with respect to content in Study 1. Although this influence of real-life knowledge worked against us, the effect was still highly robust.

To circumvent any problems of control in Study 2, we created a new set of stimuli. These “realistic” artificial diseases, although fabricated, were based almost entirely on the real biological workings of internal human systems. For instance, we described a disease of the liver that does not exist but could plausibly exist, given the actual normal workings of the human liver, with only a few minor adjustments. Using these artificial diseases, we were able to create “symptoms” at a lower, more biochemical level than the symptoms in the Study 1 stimuli. For instance, the four symptoms at the end of the two causal chains in the liver disease were four different plasma proteins normally manufactured by the liver. This enabled us to switch terminal effect symptoms back and forth between causal chains for complete counterbalancing of the stimuli, without encountering any problems in plausibility. Thus, whatever effects we might find in the participants’ responses could be attributed solely to use of the causal structure that they were given.

Method

Participants. Twenty-four Yale University undergraduates participated in this study in exchange for either partial fulfillment of the requirements for an introductory psychology course or monetary compensation at the standard rate of \$7 per hour.

Materials and Design. Three sets of stimuli were created, based on artificial diseases documenting specific “pathologies” of the brain, liver, and bone marrow, respectively. The materials were developed exactly as in Study 1, except that the symptoms were counterbalanced. To accomplish this, we invented diseases with symptoms that could readily be exchanged between causal chains,

focusing on a lower, more biochemical level of symptomatology that could be detected via analysis of blood or saliva samples. With these stimuli, two, instead of one, causal summary diagrams were developed for each disease. For instance, in the liver disease, we specified that damage to the liver tissue launches two separate causal pathways that each end in the reduced synthesis of two proteins (four proteins in all: albumin, fibrinogen, prothrombin, and globulin). For one causal summary diagram, the diversity-based answer would be the patient with reduced levels of albumin and globulin. For the other causal summary diagram, the diversity-based answer would be the other patient, who had reduced levels of albumin and fibrinogen. As in Study 1, four separate diagrams were constructed from each causal summary diagram for the participants’ task (see Table 1 for a complete description of the stimuli components used in Study 2).

Procedure. The procedure was identical to that used in Study 1 (note that there was no control condition; all the participants saw causal diagrams). No participant ever saw both sets of causal diagrams for any one disease.

Results and Discussion

The data were coded such that an assignment of “1” denoted an answer based on the diversity of symptoms, “0” denoted an answer based on causal proximity of symptoms, and “0.5” denoted that the participant indicated that he or she was unable to choose. The participants made the choice of “0.5” 6.9% of the time. As in Study 1, the data were collapsed across diseases for the analysis. A one-sample *t* test against a test value of 0.5 revealed that overall, diversity-based reasoning occurred more frequently than chance [$M = .771$, or 77.1%, of the responses, $SD = .329$; $t(23) = 4.03$, $p = .001$]. Broken down by disease, the results were in the same direction for all three items (see Figure 4).

Thus, once again, a clear within-category diversity-based effect was demonstrated. Because the materials in

Table 1
Individual Components of the “Disease” Stimuli Used in Studies 2 and 3

| Components | Disease | | |
|-----------------------|---|---|---|
| | Endocrinal Brain Encephalitis | Liver Hetroxis | Ellis-Jensen Disease |
| Root cause | brain encephalitis causes excess stimulation of certain endocrine tissues | damage to liver tissue interferes with specific functions | bone marrow produces fewer main stem cells |
| Intermediate causes 1 | mucosa of duodenum (endocrine tissue) is stimulated; alpha-islet cells in the mucosa synthesize polypeptides; activating enzyme excises portions of the polypeptides, creating hormones | liver holds fewer amino acids in storage; fewer amino acids undergo deamination; fewer deaminized compounds available for plasma protein synthesis | main stem cells produce fewer myeloid stem cells; myeloid stem cells produce fewer megakaryocytes; megakaryocytes can manufacture only limited numbers of platelets |
| Terminal effects 1 | | | |
| 1a | MSH and ANF hormones secreted | drop in prothrombin and globulin production | beta- and kappa-platelet production drops |
| 1b | MSH and CKN hormones secreted | drop in prothrombin and fibrinogen production | beta- and gamma-platelet production drops |
| Intermediate causes 2 | anterior pituitary (endocrine tissue) is stimulated; this activates neurosecretory cells which secrete a releasing factor; releasing factor stimulates secretion of hormones | liver converts glucose to glycogen more slowly; not enough glycogen to produce lipid intermediates; fewer lipid intermediates available for lipoprotein synthesis | main stem cells produce fewer lymphoid stem cells; lymphoid stem cells produce fewer lymphocytes; lymphocytes can manufacture only limited numbers of platelets |
| Terminal effects 2 | | | |
| 2a | CKN and THD hormones secreted | drop in fibrinogen and albumin production | alpha- and gamma-platelet production drops |
| 2b | ANF and THD hormones secreted | drop in globulin and albumin production | alpha- and kappa-platelet production drops |
| Patient I | analysis of a saliva sample indicates the presence of excess MSH and ANF hormones | analysis of a blood sample indicates unusually low levels of albumin and globulin | analysis of a blood sample revealed abnormally low concentrations of beta- and gamma-platelets |
| Patient II | analysis of a saliva sample indicates the presence of excess MSH and CKN hormones | analysis of a blood sample indicates unusually low levels of albumin and fibrinogen | analysis of a blood sample revealed abnormally low concentrations of beta- and kappa-platelets |

Note—These stimuli were presented to the participants in diagram format, as described in Studies 1–3. Terminal effects 1a and 2a were presented to half the participants in each study, and Terminal effects 1b and 2b were presented to the other half. The order of Patients I and II was counter-balanced between participants.

Study 2 were fully counterbalanced, the diversity-based response must be attributed to use of the specific causal structures that the participants were given.

STUDY 3

Potential Causes of the Diversity Effect

Study 2 demonstrated a strong diversity-based within-category reasoning effect in a perfectly balanced set of materials. Furthermore, any possible problems or confounds of prior knowledge in Study 1 were effectively eliminated by the design of Study 2. However, it is not clear from these studies what, specifically, drove the effect. In this study, we attempted to address two separate issues: (1) whether the effect was driven by how the participants interpreted the missing symptoms in each patient, and (2) whether the effect was driven by the participants' having to make a forced choice between two patients.

Did the participants assume in the first two studies that the symptoms not stated as being present in a patient were definitely absent? If so, the participants in those studies decided on the diversity-based response under the assumption that two, and exactly two, symptoms in

each patient were present. This would suggest that diversity per se affected the participants' decisions beyond the mere number of symptoms present.

On the other hand, were the participants instead assuming that the symptoms not stated as present were simply unknown? If this case is true, the participants might have chosen the diversity-based response for another reason entirely. In the case of the diverse patient, knowing that at least one symptom was present from each causal chain could have led them to infer that the missing symptom from each causal chain was also very likely to be present. This is because the presence of even one terminal symptom suggested that at least the first four nodes of one causal chain may have occurred, perhaps raising the perception of the likelihood that the other terminal symptom may also have occurred. Thus, if the participants thought the missing symptoms were unknown, the diverse patient may have been treated as having all four symptoms. In the case of the proximal patient, the participants might have reasoned that the presence of those two symptoms alone did not really indicate that the other two symptoms, located in an entirely different causal chain, were likely to be present. Consequently, the proximal patient may have been treated as

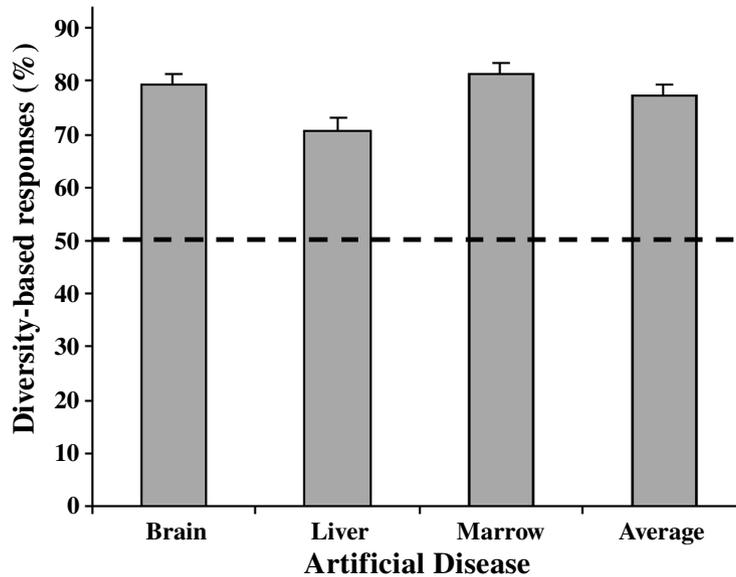


Figure 4. Percentages of diversity-based reasoning in Study 2. The horizontal dotted line denotes what responses would be at chance levels. Both overall and broken down by disease, diversity-based reasoning occurred significantly more frequently than chance. Error bars indicate standard errors.

having only the two explicitly present symptoms. To summarize, if people were interpreting the missing symptoms as being unknown rather than explicitly absent, they might have been basing their answers on the inferred number of symptoms, rather than on the diversity of symptoms per se.

To distinguish between these two possibilities of interpretation by our participants, a new between-subjects condition was added to the materials used in Study 2. Specifically, half of the participants were explicitly told that the missing symptoms were known to be absent, and half of the participants were told that the missing symptoms were unknown owing to a delay in receiving those particular laboratory test results. We were thereby able to investigate whether the participants' choice of the diverse response was affected by how they interpreted the nature of the missing symptoms in each patient.

A second issue left unaddressed by the first two studies is whether the diversity-based within-category effect is dependent on participants' having to make a forced-choice comparison. Perhaps the participants only make diversity-based answers when they feel that they are being forced to pick one response over the other. If true, this possibility would not change the finding that the participants did indeed choose that response reliably in Studies 1 and 2. However, we were interested in seeing whether the diversity effect was broader in scope than a forced-choice situation. Thus, in Study 3, we examined whether people make diversity-based responses even

when not asked to pick one patient over the other. We implemented this manipulation by changing the forced-choice format to an individual rating scale for each patient. In this way, we were able to measure the responses the participants made on an absolute scale for each patient in turn. The rating scale dependent measure would also enable us to collect data that are continuous rather than categorical in nature, allowing us to run a complete ANOVA.

Method

Participants. Twenty-four McMaster University undergraduates participated in this study in exchange for either partial fulfillment of the requirements for an introductory psychology course or monetary compensation at the standard rate of \$7 per hour.

Materials and Procedure. The materials and procedure were identical to those used in Study 2, with several modifications. First, a sentence was added to each patient description specifying whether the state of the missing symptoms was known or unknown. Half the participants were told unequivocally that the missing symptoms were not present in the patient. The other half was told that the state of the missing symptoms was unknown at this time, because the laboratory results needed to confirm their presence or absence were not yet available.

Second, the dependent measure question was altered such that the participants were simply asked for each of the two patients, "How likely is it that [Patient X] has [Disease Y]?" The participants were asked to give their answers on a scale of 0–100, where 0 = a 0% chance and 100 = a 100% chance. As in Studies 1 and 2, patient descriptions and the dependent measure questions for both the diverse and the proximal patients were printed on the same page. At the top of the page, the participants were instructed to read both of

the following problems before giving their answers. This was added to ensure that the participants' responses did not go to floor or to ceiling simply because of their response to whichever patient they rated initially.

Finally, the order in which the participants saw the causal diagrams for each disease was pseudorandomized, rather than having two preset pseudorandomized orders as a between-subjects condition. Again, these orderings were determined randomly, except that diagrams containing symptoms that were products of the same causal chain were not placed consecutively.

Results and Discussion

A 2 (patient type: diverse or proximal) \times 2 (missing type: explicit or unknown) \times 3 (disease: brain, liver, or marrow) ANOVA was conducted. The main effect of patient type was significant, such that the diverse patient ($M = 61.7\%$, $SD = 25.4\%$) was given a higher likelihood of having the disorder than was the proximal patient [$M = 49.1\%$, $SD = 25.5\%$; $F(1,22) = 11.82$, $MS_e = 483.42$, $p = .002$]. Thus, these results replicated the diversity effect found in Studies 1 and 2.

These results are further supported by a marginally significant interaction of missing type and patient type [$F(1,22) = 3.69$, $MS_e = 483.42$, $p < .07$]. Although the direction of the results was the same in both the explicit and the unknown conditions, the diversity effect was significant in the explicit condition, but not in the unknown condition (see Figure 5). Broken down by condition, the participants in the explicit condition showed a significant diversity effect (mean rating of 65.2%, $SD = 24.6\%$) for the diverse patient versus 45.6% ($SD = 23.1\%$) for

the proximal patient [$t(11) = 3.61$, $p = .004$], whereas the participants in the unknown condition did not (mean rating of 58.2%, $SD = 35.2\%$) for the diverse patient versus 52.6% ($SD = 34.6\%$) for the proximal patient [$t(11) = 1.13$, $p = .3$]. This finding suggests that the diversity effect found in Studies 1 and 2 was not driven by the assumption that the status of the missing symptoms was unknown. It is not clear why unknown status in the missing symptoms would diminish preference for diversity; one possibility is that uncertainty in the cover story may precipitate a tendency toward greater overall caution when a judgment is made. Future studies may determine whether this is the case. No other main effects or interactions reached significance (all $ps > .3$).

Broken down by disease, the diversity effect appeared in all items to a similar degree. For the liver disease, disease likelihood ratings for diverse patients ($M = 62.1\%$, $SD = 28.7\%$) were significantly higher than ratings for proximal patients [$M = 47.9\%$, $SD = 27.7\%$; $t(23) = 2.89$, $p < .01$]. Similarly, for the marrow disease, ratings for diverse patients ($M = 60.9\%$, $SD = 28.7\%$) were higher than ratings for proximal patients [$M = 48.3\%$, $SD = 30.7\%$; $t(23) = 2.37$, $p < .03$]. Finally, for the brain disease, ratings for the diverse patient ($M = 62.1\%$, $SD = 27.3$) were nearly significantly higher than those for the proximal patient [$M = 51.0\%$, $SD = 24.5\%$; $t(23) = 2.37$, $p < .06$]. Thus, the main effect of patient type shown by the ANOVA was not driven by any single item.

Analyses were also run to compare these results with those of Study 2. Responses from the present study were

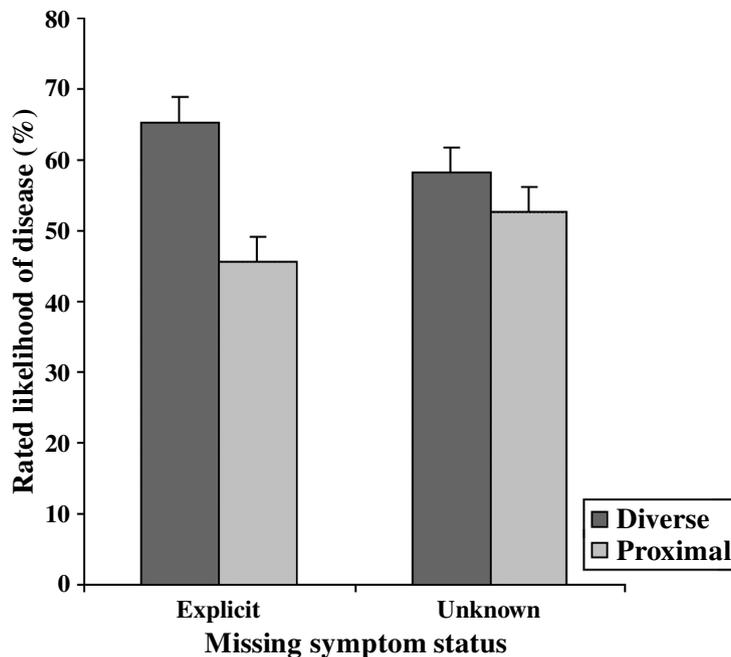


Figure 5. The participants' ratings of disease likelihood for the items in Study 3, broken down by missing symptom status (explicitly absent vs. unknown status). Error bars indicate standard errors.

recoded to convert the present study data to the same format as that for the Study 2 data. Specifically, an assignment of “1” was given whenever the participant gave a higher rating to the diverse patient than to the proximal patient, an assignment of “0” was given for the reverse pattern, and an assignment of “0.5” was given when the participant gave equal ratings to both patients. A one-sample *t* test of the data, collapsed over disease, was conducted against a test value of 0.5. The results showed that the diversity-based response, even coded as all-or-none, was chosen significantly above chance [$M = .646$, or 64.6%, of the responses, $SD = .292$; $t(23) = 2.45$, $p = .02$]. A two-tailed independent-samples *t* test was also conducted to determine whether the forced-choice question used in Study 2 ($M = .771$, or 77.1%, of the responses, $SD = .329$) elicited more diversity-based responses than did the independent rating scales used in Study 3. There was no significant difference between the two studies [$t(46) = 1.39$, $p < .2$].

Summary. Study 3 gave additional evidence for the within-category diversity-based effect demonstrated in Studies 1 and 2. This time, the effect was replicated with a rating scales dependent measure that did not force the participants to choose between patients, testifying to the strength of the diversity effect. The effect was significant when the absence of symptoms was made explicit, but not when it was left uncertain. This suggests that the robust diversity effect found in all three studies was likely due to diversity per se, rather than to the total number of symptoms inferred to be present.

GENERAL DISCUSSION

In these studies, we have demonstrated a diversity effect for chains of causal processes that is analogous to the diversity effect found for reasoning about hierarchically embedded categories. In cases with a single cause, one can envision the immediate consequences of that cause as branching out into a hierarchical structure. Those consequences, if known to be causes, dominate subsequent causes on lower branches in the tree.

For several possible reasons, people see consequences that cluster in one area of the tree as less diagnostic of the initial cause than are those that are more diversely represented. One reason may be a greater ease in envisioning alternative root causes if the presenting patient's symptoms are tightly clustered in one area of the causal tree. Having diverse symptoms in the hierarchy, on the other hand, would be useful in decision making, because they would help rule out alternative causes. One might reason that having two symptoms from different branches in the hierarchy would make it quite unlikely that any other cause could be responsible for two such diverse results. This resulting dearth of plausible alternatives strengthens the explanatory coherence of the initial-cause explanation (Read & Marcus-Newhall, 1993). This ruling out of alternative causes may be specific to

causal chains and not applicable to the diversity effect for categories.

Categorical and causal diversity may both sometimes involve the interpretation of conjunctive features. That is, one approach to understanding causal diversity may involve decomposing the original cause into two component causes right from the start. To use a contrived example, a physician might mentally divide radiation exposure into two component causes, the burning of superficial skin layers and the destruction of white blood cells. The physician might further assume that the likely consequences of superficial skin layer burning (i.e., permanent skin darkening, scarring of the skin) are independent of the likely consequences of white blood cell destruction (i.e., pneumonia, yeast infection in the mouth). To be more confident that a given patient has radiation sickness, the physician might favor evidence that points back to both of the original causes, rather than to just one. In the radiation sickness case, this may involve the use of a feature-checking strategy, in which features are assumed to have causal consequences. Thus, in cases in which the original cause does not result in two other causes but, rather, is itself decomposable into two causes, the inductive strategy may be to favor evidence that supports the presence of those two original causes. This sort of diversity-based reasoning also seems to have a parallel in category-based induction (i.e., the category of animals could be immediately parsed into multiple subcategories, and so on).

A final question about mechanisms for causal diversity concerns two different ways in which proximity might be understood. One might see two terminal effects as similar because they are causally interdependent. If having to buy bigger sized clothes and weighing more on the bathroom scale are seen as nondiverse, it may be because we cannot imagine one's occurring without the other. Hence, such pairs of effects are not independently informative. That sort of proximity effect seems independent of the sort of hierarchical branching structure discussed until now. Alternatively, one might think that two terminal effects occur independently but share a common local superordinate cause. In those cases, which are presumably the same as those used in the present studies, changing the causal branch structure can have dramatic effects.

The cases used in the present studies are idealized in several respects. The branching structures are symmetrical and go to the same depth throughout, and there is a clear single cause. In most real-world cases, the hierarchies will be messier and may often have links to more than one cause. Such factors might make computations of effects more difficult, but we believe that the basic diversity effect would hold in these cases as well, given the robustness of our findings. For any system with chains of causes and consequences, it seems likely that some variant of the diversity effect is likely to hold. In future studies, it will be important to explore which of the possible

mechanisms for the causal diversity effect is most influential in different contexts.

A related issue concerns how implicit the knowledge of the hierarchical structure can be. Although these studies did not directly address that point, comments by some participants in debriefing suggest that few, if any, envisioned hierarchical tree structures of the sort shown in Figure 2. They did, of course, have the explicit knowledge of more local causal chains supplied to them in the study. It would be useful to explore this issue more directly through implicit learning paradigms in which even local causal chains were not explicitly taught.

One of the most intriguing features of the causal diversity effect is that it could conceivably be used to explore people's intuitive causal beliefs in different domains. For example, imagine that a person is gaining expertise in a domain and that the acquisition of expertise fosters the construction of clearer theories about hierarchical causal relations. In this case, we would expect that gaining expertise would also give rise to increasingly powerful diversity effects. Similarly, dramatic conceptual changes from one set of causal beliefs to a different one might well result in a shift in focus of the diversity effect, as was demonstrated in Study 2. Moreover, end state effects that are diverse according to one person may be just the opposite for another who has a different causal hierarchy. These issues might be most dramatically observed in developmental studies, in which changes in the magnitude and location of the effect could be used to make inferences about underlying changes in causal knowledge with age.

Although disease diagnosis may be the most obvious case, we have argued that diagnostic decisions are ubiquitous in our daily lives and are seen in reasoning about our cars, our friendships, and ailing economies. How might the diversity effect vary across domains, given differences in the causal structures associated with those domains? For example, in some mechanical systems, the presence or absence of a symptom may be understood as more deterministic and less probabilistic than in the domain of human personality. Similarly, some domains may tend to have more deeply structured hierarchies than others. The diversity of the diversity effect across domains remains an important question to investigate.

REFERENCES

- AHN, W., & KALISH, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199-225). Cambridge, MA: MIT Press.
- AHN, W., KIM, N. S., LASSALINE, M. E., & DENNIS, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, **41**, 361-416.
- CAREY, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Plenum.
- 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.
- GUTHEIL, G., VERA, A., & KEIL, F. C. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition*, **66**, 33-49.
- HEIT, E., & HAHN, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology*, **43**, 243-273.
- INAGAKI, K., & HATANO, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychology Press.
- JENSEN, F. V. (1996). *An introduction to Bayesian networks*. New York: Springer-Verlag.
- LÓPEZ, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition*, **23**, 374-382.
- LÓPEZ, A., GELMAN, S. A., GUTHEIL, G., & SMITH, E. E. (1992). The development of category-based induction. *Child Development*, **63**, 1070-1090.
- MAGNUS, D. (1996). Heuristics and biases in evolutionary biology. *Biology & Philosophy*, **12**, 21-38.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- READ, S. J., & MARCUS-NEWHALL, A. R. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality & Social Psychology*, **65**, 429-447.
- 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.
- RUSE, M. (1998). Answering the creationists: Where they go wrong—and what they're afraid of. *Free Inquiry*, **18**, 28-32.
- SPELLMAN, B. A., LÓPEZ, A., & SMITH, E. E. (1999). Hypothesis testing: Strategy selection for generalising versus limiting hypotheses. *Thinking & Reasoning*, **5**, 67-91.
- THAGARD, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.
- WHEWELL, W. (1847). *The philosophy of the inductive sciences*. London: Frank Cass.
- WHEWELL, W. (1999). *The philosophy of the inductive sciences, founded upon their history*. Bristol: Thoemmes Press. (Original work published 1840)

(Manuscript received November 2, 2001;
revision accepted for publication August 21, 2002.)