American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/20/$12.00

2020, Vol. 46, No. 2, 296–315
http://dx.doi.org/10.1037/xlm0000709

The Contrast Diversity Effect: Increasing the Diversity of Contrast Examples Increases Generalization From a Single Item

David A. Kalkstein
Stanford University

David A. Bosch
New York University

Tali Kleiman
The Hebrew University of Jerusalem

In five experiments, we established and explored the contrast diversity effect—the effect of diversity of negative evidence on inductive inferences drawn from a single observation of a target exemplar. In Experiments 1 through 3, we show that increasing the diversity of negative evidence in a contrasting category led people to infer that a target exemplar corresponded to a higher level category and led to greater generalization of a novel property associated with the target. Further, we demonstrated two boundary conditions in which the effect only occurred when the negative evidence was consistent with a higher level category that both united the contrast exemplars and distinguished them from the target (Experiment 4) and when the negative evidence and the target shared an obvious parent category (Experiment 5). Taken together, these findings demonstrate that increasing the diversity of negative evidence alone increases generalization from a target so long as the negative evidence is drawn from a single contrast category that excludes, but shares a common parent with, the target. Implications for general theories of induction are discussed.

Keywords: contrast categories, diversity, generalization, induction, negative evidence

The human propensity for inductive reasoning—making inferences that are based on, but go beyond, available data—has long been recognized by psychologists as one of the most adaptive aspects of human cognition (Nisbett, Krantz, Jepson, & Kunda, 1983; Osherson, Smith, Wilkie, López, & Shafir, 1990; Shipley, 1993; Xu & Garcia, 2008). These inferences allow people to form expectations about future events and states of the world, relying on past experience to provide a basis for prediction, decision making, and action. For example, if a child tries an unpleasantly bitter Brussels sprout and subsequently forms the belief that all Brussels sprouts are bitter, she has made an inductive inference. This inference may in turn guide her decision making and behavior in the direction of avoiding Brussels sprouts altogether. This is an example of a category-based induction, as the induction of the property *bitter* was generalized from the single observation to all items belonging to the category *Brussels sprouts* (see Osherson et al., 1990; Rips, 1975). An even broader category-based induction is also possible: from the observation of a bitter Brussels sprout, the child may infer that *all vegetables* are bitter.

To explain inductive reasoning, it is crucial to understand both what promotes and constrains generalization from limited observations. Generalization refers to the range, or breadth, of objects (or categories) that people infer possess a given property[1] observed in some target. As the example above demonstrates, categories can provide one natural source of constraint on inductive inferences (Murphy, 2010). Higher level categories (e.g., vegetables) are composed of a broader array of objects and have greater diversity among their members than do lower level categories (e.g., Brussels sprouts). Thus, one way that greater generalization manifests is through category-based inductions made along higher level (i.e., more superordinate) category lines.

Previous work has shown that a variety of factors—ranging from diversity of the objects observed to possess a given property to their prototypicality—influence category-based inductions (e.g., Osherson et al., 1990; Rips, 1975). For the most part, this research has explored how characteristics of objects shown to possess a given property impact people's generalization of that property. In

This article was published Online First May 13, 2019.

David A. Kalkstein, Department of Psychology, Stanford University; David A. Bosch, Department of Psychology, New York University; Tali Kleiman, Department of Psychology, The Hebrew University of Jerusalem.

We thank Gregory L. Murphy for providing helpful comments on an earlier version of this article. Interested readers may access the original stimuli that were shown to participants in this and all remaining experiments at https://osf.io/twery/.

Correspondence concerning this article should be addressed to David A. Kalkstein, Department of Psychology, Stanford University, Jordan Hall Building 420, 450 Serra Mall, Stanford, CA 94305. E-mail: dave.kalkstein@gmail.com

[1] To define our terms, we use the term dimension to refer to a variable-like set of qualities or characteristics (usually mutually exclusive) that contains within it varying properties (or values). We use property and value somewhat interchangeably to refer to specific attributes of an object or category. For example, taste is a dimension whose values can vary across objects or categories—within the dimension of taste, a Brussels sprout may possess the property of being bitter while an apple may possess the property of being sweet.

contrast, relatively little work has explored how characteristics of objects shown to *not* possess a given property impact people's inductive reasoning about that property. In this paper, we introduce the contrast diversity effect to examine the impact that diversity in observations of objects that do *not* possess a given property (i.e., diversity of negative evidence) has on the category level that people use in their inductive inferences, and thus on how broadly people generalize a novel property shown in a single target observation.

Diversity in Induction

Past research on inductive reasoning has focused on how the diversity of evidence within a reasoning task influences the conclusions that people draw from that evidence (see Hayes, Heit, & Swendsen, 2010; Heit, 2000 for reviews). This work has found that when encountering a novel set of items that all possess a given property (e.g., bitterness), people tend to infer that the property is characteristic of the lowest level known category that is consistent with the observed set of exemplars. As a result, people tend to limit their generalizations of that property to members of that lowest level category (e.g., López, 1995; Osherson et al., 1990; Xu & Tenenbaum, 2007a, 2007b). Because a less diverse set of exemplars can be covered by a lower level category, this finding implies that less diversity within a set of exemplars shown to possess a specific property will lead to more narrow generalization of that property (i.e., generalization across a lower level category) and that greater diversity will lead to broader generalization of that property (i.e., generalization across a higher level category).

For example, if a person is shown three different apples and told that each has some property X, he would likely limit his generalization of the property X to members of the basic level category *apples* because it is the lowest level category that covers the diversity of the observed exemplars. If given a more diverse set of examples, such as an apple, a banana, and a strawberry and told that each has property X, he would likely extend his generalization of property X to the more superordinate category *fruits* because it is now the lowest level category that covers the diversity of the observed exemplars. Taken together, previous research exploring the role of diversity in inductive reasoning suggests that greater diversity of evidence leads to greater generalization.

Importantly, however, past research has focused on how far people generalize a given property based on the diversity of *positive evidence*—observations of objects shown to possess the relevant property (e.g., Osherson et al., 1990; Rips, 1975; Sloman, 1993). In the present research, we ask the novel question of how diversity of *negative evidence* influences inductive reasoning.

Negative Evidence

Whereas observations of target objects that possess the to-be-generalized property constitute positive evidence, observations of objects that do NOT possess the to-be-generalized property constitute negative evidence (Heussen, Voorspoels, Verheyen, Storms, & Hampton, 2011). For example, if the property in question is the bitter taste observed in a Brussels sprout, the observation that broccoli is also bitter is an example of positive evidence, whereas the observation that apples are sweet (which implies NOT bitter) is an example of negative evidence (see Kalish & Lawson, 2007). A central finding from research on negative evidence is that given an observation of a target object possessing a given property, the addition of negative evidence can actually increase generalization when the positive and negative observations are drawn from contrasting categories—a set of mutually exclusive subcategories that share a common parent (Heussen et al., 2011; Kalish & Lawson, 2007; Voorspoels, Navarro, Perfors, Ransom, & Storms, 2015). For example, given the premise that music by Mozart elicits alpha waves, people are more likely to generalize the property of eliciting alpha waves to the music of Bach when they are given the additional negative evidence that the music of AC/DC *does not* elicit alpha waves than when they are given no additional information (Heussen et al., 2011).

In general, although much is known about how characteristics of positive evidence (e.g., its diversity and prototypicality) influence inductive reasoning and generalization (e.g., Heit, 2000; Osherson et al., 1990; Rips, 1975; Sloman, 1993), comparatively little is known about how these same characteristics impact induction when displayed in negative evidence. As a result, it is not well understood how, or even whether, principles of inductive reasoning from positive evidence apply to negative evidence. In the current research, we begin to fill this gap in the literature by exploring the impact that one prominent characteristic of negative evidence—its diversity—has on inductive reasoning.

The Present Research—The Contrast Diversity Effect

In the present research, we integrate research on diversity and negative evidence in induction to introduce the contrast diversity effect. Specifically, we propose that the extent of generalization from a single target exemplar will change based on the diversity of negative evidence alone. To illustrate, imagine a simple game wherein different items are presented to you one at a time and some items give you money whereas others do not. Imagine that on one turn an apple shows up and you win money, but on each of the next three turns a carrot shows up and you do not win. The conclusion here may be that apples give money whereas carrots do not. But what if, instead, you see that you win when an apple appears but do not win when broccoli, lettuce, and a carrot show up? Perhaps now the conclusion would be that fruits give money while vegetables do not. In this example, the reward associated with the apple is the to-be-generalized property. The *negative evidence* is represented by the set of objects drawn from the contrasting category *vegetables* that all had a reward value of zero, and the *target object* (positive evidence) is represented by the item that gave a reward. Note that in both scenarios the target object was represented by the same single item—an apple—and that only the negative evidence changed.

As the example above intuitively suggests, we hypothesize that increasing the diversity of negative evidence should lead to increased generalization of a novel property introduced in a single piece of positive evidence (that is, when the negative evidence can be covered by a single contrasting category that excludes the target). Such an effect, we argue, will manifest through the level of the category that is used in category-based inductions of the target property.

Level-Matching

Our proposal of the contrast diversity effect and the predictions stemming from it draws on past research demonstrating that people possess a cognitive bias to infer that contrasting sets are consistent in their structure and detail (e.g., Billman & Dávila, 2001; Shipley & Kuhn, 1983). When prompted to generate a target category from a given contrast, people are influenced by the level of the provided contrast category (Malt & Johnson, 1992; Markman & Wisniewski, 1997; Verbeemen, Vanoverberghe, Storms, & Ruts, 2001). For example, when asked to generate a target category from the contrasting category of *fruit*, an overwhelming majority of participants produced *vegetables* as the appropriate target category, while no one produced a more subordinate category (e.g., *carrots*; Storms, De Boeck, & Ruts, 2001). These observations support the intuition that when a set of negative exemplars is represented as a higher level contrast category, people will raise their conceptual representation of the target category to what they deem to be an equivalent level of abstraction despite having no new information about the target.

Applying this logic to the current research, we argue that increasing the diversity of the negative exemplars increases the inferred level of the contrast category (e.g., Osherson et al., 1990; Xu & Tenenbaum, 2007a, 2007b), and that the level at which the contrast category is represented influences the category level that the target object is inferred to represent (e.g., Billman & Dávila, 2001; Malt & Johnson, 1992; Markman & Wisniewski, 1997; Verbeemen et al., 2001). The level inferred for the target category then guides inductive generalization by specifying the range of objects to which an observed property may extend.

Heuristically, this process can be thought of as matching the taxonomic level of the target category with that implied by the contrast. So, if a person infers that the contrast category is carrots, the apparent commensurate category level for the target category may be apples. However, if the contrast category is represented at a higher level, such as vegetables, the apparent commensurate category level for the target category may be fruit.

Relation to Previous Accounts of Inductive Reasoning

Testing the contrast diversity effect and our proposed level-matching account—that people use the category level implied by the negative evidence to induce a target category at an equivalent level of abstraction—provides an opportunity to explore how patterns in inductive reasoning from mixed observations (i.e., both positive and negative evidence) compare with previously documented patterns in inductive reasoning from positive evidence alone. In what follows, we detail some specific ways in which the contrast diversity effect is distinct from previous accounts of induction and discuss its implications for general theories of induction.

Inductive reasoning from negative versus positive evidence. If observed, the contrast diversity effect would suggest that empirically established principles of induction from positive evidence do not necessarily, or directly, apply to inductive reasoning that includes negative evidence as well. For example, the predicted contrast diversity effect suggests that the *Size Principle*—which has been consistently demonstrated in research on people's reasoning from positive evidence—may not hold when people reason from mixed evidence. According to the Size Principle, more narrow extensions (i.e., less generalization) are more likely than broader extensions (i.e., greater generalization) when both are consistent with a given set of examples and, further, the preference for narrower generalizations increases exponentially with the number of examples observed (see Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007a). This explains why people would be more likely to limit their generalization of Property X only to apples after observing three different apples all containing Property X than after observing just one apple containing Property X.

In predicting the contrast diversity effect, we expect that increasing the diversity of contrasting negative exemplars will increase generalization from a target exemplar to a broader category—even when a narrower category is consistent with the data. Moreover, we expect people will display a tendency for broader generalization even when increasing the diversity of negative evidence means including more unique observations (e.g., contrasting an apple with a carrot, lettuce, and broccoli vs. contrasting an apple with just a carrot). This prediction, which we test below, directly contradicts the second clause in the size principle that increasing the number of exemplars seen increases the preference for more narrow generalization.

Category-based induction. The research presented below may also shed light on the debate between theories that posit a central role for categories in inductive reasoning (e.g., Coley, Medin, & Atran, 1997; Medin, Coley, Storms, & Hayes, 2003; Osherson et al., 1990; Proffitt, Coley, & Medin, 2000) and those which claim that it is unnecessary to assume that people rely on categories when making inductive inferences (e.g., Sloman, 1993). Our level-matching account assumes that people rely on prior knowledge of hierarchical categories in inductive reasoning from negative evidence and runs counter to a feature-based account that relies solely on judgments of similarity between presented evidence and possible conclusions to guide inductive reasoning.

A pure similarity account would hold that the likelihood of generalizing a given property to a novel object is directly related to the similarity between the observed positive evidence and the novel object and is inversely related to the similarity between the negative evidence and the novel object (e.g., Osherson et al., 1990). This implies that more diverse positive observations would lead to greater generalization because the increased diversity in positive exemplars increases the likelihood that a novel target will be similar (i.e., have similar features) to an already observed positive exemplar. On the other hand, and pertinent to the present research, this would imply that increasing the diversity of negative evidence should decrease generalization since it increases the likelihood that a novel target will be similar to an observed exemplar that does NOT possess the given property. In contrast, our proposed level-matching account uniquely predicts that increasing the scope of objects shown to NOT possess a target property should also increase the scope of objects that people infer DO possess the target property. So whereas a feature-based similarity model of induction would be hard-pressed to understand why increasing the breadth of things observed to *not possess* a given property would lead people to infer a greater breadth of objects *do possess* that property, such a finding would easily be accounted for by a process of level-matching across negative and positive evidence.

Unique Predictions

If the contrast diversity effect does indeed rely on category-based induction and a process of level-matching, then it follows that the extent of generalization from a target observation will be based on which category level is made salient by the negative evidence (see also Medin et al., 2003; Voorspoels et al., 2015). Hence, increasing the diversity of negative evidence should only increase generalization when the negative evidence suggests a contrast category that highlights the target object's inclusion in a higher level category. Supporting this category-based account, we predict that the contrast diversity effect will occur only when the negative evidence can be covered by a single contrasting category that is distinct from the positive evidence but shares an overarching parent category.

This prediction suggests at least two major boundary conditions. First, increasing the diversity of the negative evidence should increase generalization from a target up until the negative evidence becomes so diverse that the lowest level category that covers it also includes the target object. At this point, increasing the diversity of negative evidence will no longer make salient a higher level category for the target. For example, relative to a case where an apple is contrasted with a carrot, we would not expect that contrasting an apple with a carrot, salmon, and cheese would increase generalization of a property displayed by the apple since this latter set of negative evidence can only be covered by a category (e.g., food) that also includes the target and thus does not highlight the unique superordinate category of the target (e.g., fruit).

If observed, this nonmonotonic relationship between diversity of negative evidence and generalization would represent another way in which the level-matching account of the contrast diversity effect is distinct from direct application of previously established principles of induction from positive evidence. Within the domain of positive evidence, previous research has consistently documented a pattern—termed the Monotonicity Principle—wherein increasing the diversity of observed positive evidence monotonically increases the scope of generalization (e.g., Osherson et al., 1990; Sloman, 1993). In contrast, because we expect the contrast diversity effect to occur through a process of level-matching, we do not expect increasing diversity of negative evidence to monotonically increase generalization from the target evidence. Instead, we expect that increasing the diversity of negative evidence to the point where it suggests a higher level category that also includes the target will lead to decreased generalization—hence, a nonmonotonic relationship.

A second predicted boundary condition that follows from our level-matching account is that the contrast diversity effect should only occur when the negative and positive evidence are drawn from contrasting subcategories that share a common parent. Although shared representational structure of subcategories that have a common parent allows people to intuit what would be equivalent levels of abstraction for negative and positive evidence when they are drawn from subcategories within a given hierarchy (Malt & Johnson, 1992; Markman & Wisniewski, 1997; Storms et al., 2001; Verbeemen et al., 2001), people likely lack clear intuitions about equivalent levels of abstraction for subcategories across divergent parent categories. In other words, it is much more intuitive to match the hierarchical level of fruits and vegetables, or apples and carrots, than it is to match the level of fruits and tools, or apples and hammers. As a result, we predict that the degree of diversity in negative evidence will have less, or no, effect on generalization when the negative and positive evidence are drawn from categories that do not share an obvious parent.

If observed, this boundary condition would further cast doubt on a process of feature-based similarity judgments as the central guiding factor in inductive reasoning from negative evidence. A pure similarity account would likely predict that increasing the similarity of the negative evidence to possible conclusion targets would decrease generalization. In contrast, we predict that increasing the diversity of negative evidence will increase generalization *precisely when* that negative evidence is drawn from a superordinate category that is closely related to the target observation—namely, a contrast category that shares a common parent. Because contrast categories are more similar to each other than are categories that do not share a common parent, this boundary condition amounts to the prediction that, at least in some cases, increasing the similarity of the negative evidence (i.e., objects that *do not* possess the property in question) to the positive evidence (i.e., the target possessing a given property) will actually increase generalization from the target object (see also Heussen et al., 2011; Voorspoels et al., 2015).

Summary of Current Research

With the contrast diversity effect, we aim both to demonstrate a novel effect and to document a consistent pattern of inference that may speak to larger theories of induction. We propose that people use the category level implied by the diversity of negative evidence to infer the appropriate category level for objects possessing a target property. Supporting the level-matching account of the contrast diversity effect, we argue that the contrast diversity effect will only occur when the negative evidence is drawn from a single contrasting category that is distinct from the positive evidence but shares an overarching parent category, namely, only when the negative evidence makes salient an obvious category for the target.

Across five experiments, using different stimuli and different properties, we demonstrate our proposed contrast diversity effect and document two critical boundary conditions. Experiment 1 provides initial evidence documenting our predicted contrast diversity effect. In what follows, we seek to build on this finding by demonstrating the contrast diversity effect with different categories (Experiment 2), with different properties and different degrees of diversity (Experiment 3), as well as document important boundary conditions (Experiments 4 and 5).

Experiment 1

In Experiment 1, we sought to demonstrate the basic effect of diversity of negative evidence within a contrasting category on inductions made from a single instance of a target exemplar. We hypothesized that, compared with a set of negative observations without diversity, a diverse set of negative observations drawn from a contrasting category would lead participants to represent the target category at a higher level of abstraction and to generalize the property associated with the target exemplar to a broader array of objects.

Method

Participants. Two hundred one participants (111 female; M_{age} = 38.43, SD_{age} = 12.66) were recruited online using Amazon's Mechanical Turk. The sample size in this survey was calculated using G*Power v3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to achieve 80% power to detect a small to medium effect size (d = .40). Participants were randomly assigned to either the no-diversity condition or the diverse condition. This and all experiments reported here were approved by the Institutional Review Board.

Materials. Color pictures of six different objects taken from the general category of sports equipment were used. The six objects were distributed between two contrasting categories: baseball equipment (a baseball bat, a baseball, and a baseball glove) and hockey equipment (a hockey stick, a hockey puck, and hockey skates). Depending on condition assignment, participants saw as negative evidence either one object from one category repeated three times (no diversity condition) or three different objects from one category (diverse condition). In both conditions, the target was one object from the other category.

Procedure. Participants were told that they would be playing a simple game where the goal was to win as many points as possible, and that the game consisted of two phases, a learning phase and a test phase. They were told that in the learning phase they would see different objects and learn how much each object was worth and that in the test phase they would be asked questions and make choices based on what they learned. In this and all following experiments we emulated previous work on inductive reasoning (e.g., Osherson et al., 1990; Rips, 1975; Voorspoels et al., 2015) by presenting participants with a simplified learning task that presented a limited number of learning trials.

The learning phase consisted of four trials. The order of trials was randomized. On each trial participants saw a picture of an object and the number of points that the object was worth printed below the picture (see Figure 1). Each trial was presented for three seconds, after which participants could click an arrow to continue onto the next trial. On three of the trials, the message "+0 points" was printed in red below the picture of the object. On the other trial, the message "+10 points!" was printed in green below the picture. Thus, three of the four pictured objects were nonrewarding and only one was rewarding. This distinction established a target (the rewarding object) and the contrasting negative evidence (the nonrewarding objects). The single target example was from one of the two categories (e.g., baseball equipment) while the negative examples were from the other (e.g., hockey equipment). For half of the participants, the single target example was a baseball bat and the negative examples were all examples of hockey equipment. For the other half of participants, the target example was a hockey stick and the negative examples were all examples of baseball equipment.

What participants saw as negative evidence depended on whether they were in the no-diversity condition or the diverse condition. Figure 1 depicts a representation of what each condition looked like for participants who saw a baseball bat as the target. In the no-diversity condition, the three nonrewarding trials all displayed the same picture of the same object (top row of Figure 1). In the diverse condition, the three nonrewarding trials each displayed a different object from the same contrasting category (bottom row of Figure 1).

After completing the learning phase, participants proceeded onto the test phase, where they answered questions based on what they had learned. First, participants typed in an open-ended re-

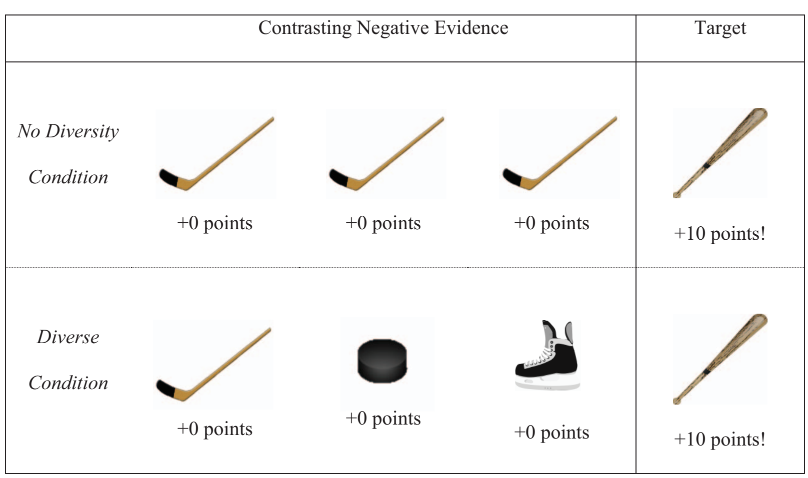

Figure 1. Depiction of the stimuli used during the learning phase of Experiment 1. Each item in the table, along with its point value, represents one trial. Trials were presented to participants one at a time on separate screens in a random order. The stimuli depicted were used for participants who were assigned to see a baseball bat as the target and hockey sticks/hockey equipment as contrasts. Other participants received a hockey stick as the target and baseball bats (no diversity condition)/baseball equipment (diverse condition) as the contrasts. The images shown here, and in all subsequent figures, are presented to give readers a sense of how the experiment appeared to participants. However, for copyright reasons, the images depicted in these figures are not the same as those shown to participants though they are highly similar. Interested readers may access the original stimuli that were shown to participants in this and all remaining experiments at https://osf.io/twery/. See the online article for the color version of this figure.

sponse to the question, “Based on what you learned, what gives you points in the game?” Next, as a direct measure of generalization, participants rated on a six-point scale how likely various baseball related objects (baseball hat, baseball jersey, and baseball cleats) and hockey related objects (hockey helmet, hockey jersey, and hockey gloves) were to give them points (1 – *extremely unlikely*; 6 – *extremely likely*). These options were embedded in a set that also included unrelated sports equipment (basketball shorts, soccer ball, football helmet, football pads). Finally, participants were asked “Which of the following descriptions best captures how you represented the rewarding stimulus in your mind?” They chose between a lower level description (e.g., baseball bat) and a higher level description (e.g., baseball equipment).

Results

We predicted that showing participants a diverse set of negative exemplars from a contrasting category would lead them to generalize from the target object to a greater extent than showing participants a single exemplar (repeated three times) as the negative observation. We tested this prediction for the three complementary measures described above. We first coded participants’ responses to the open ended question of what gave them points in the game for the level of the category that they stated gave them points. Participants who stated that a more superordinate category (i.e., “baseball equipment” or “hockey equipment”) was rewarding received a code of 1, participants who said that a more subordinate category (i.e., “baseball bat” or “hockey stick”) was rewarding received a code of 0. We coded responses that included two categories according to the highest level category provided in the answer. Anything other than these responses (e.g., “I forgot,” “Single syllable words,” “Objects”) was not coded and left as missing ($n = 28$, 14% of responses). The percentage of participants providing a codable response did not differ between conditions (84% in the diverse condition vs. 88% in the no diversity condition, $\chi^2(1, N = 201) = .71$, $p = .40$). As predicted, a greater percentage of participants stated that the higher level category (baseball equipment or hockey equipment) was what was rewarding in the diverse condition (19%) than in the no diversity condition (1%),[2] $\chi^2(1, N = 173) = 15.67$, $p < .001$.[3] There was not a significant effect of which category (baseball or hockey) the target object was from. This is true of the rest of the experiments in this paper and thus will not be discussed further.

Next, we averaged together participants’ ratings of the likelihood that an item from the target’s superordinate category would earn them points (e.g., if the participant saw a baseball bat gave them 10 points in the learning phase, this measure was the average rating of the likelihood that a baseball hat, a baseball jersey, and baseball cleats would also give them 10 points). This composite measured achieved good reliability ($\alpha = .93$). We compared this average likelihood rating across conditions and found that participants in the diverse negative evidence condition rated other items from the same superordinate category as the target item as more likely to be rewarding ($M = 4.71$, $SD = 1.52$) than participants in the no diversity condition ($M = 4.27$, $SD = 1.44$), $t(199) = 2.10$, $p = .04$, $d = .30$. Item by item analyses for this and all experiments are shown in the Appendix.

Finally, we looked at the forced choice question to see whether participants identified the target object at a higher level when presented with a variety of contrasting examples. We found a marginally significant effect in the predicted direction such that a greater proportion of participants preferred the higher level description of the target in the diverse condition (37%) than in the no-diversity condition (25%), $\chi^2(1, N = 201) = 3.53$, $p = .06$.

Discussion

The results of Experiment 1 support our hypothesis that the breadth of generalization from a single target exemplar increases when introducing diversity into the negative evidence. Specifically, participants were more likely to infer that the property associated with a single target object was characteristic of a higher level category and were more willing to extend the reward value of the target exemplar to other members of the target’s superordinate category when shown a diverse set of contrasting negative exemplars than when shown a single contrasting negative exemplar. Taken together, these findings demonstrate that people tend to generalize from a single target example more broadly when that target is presented with a more diverse set of negative evidence drawn from a contrasting category.

These results demonstrate a violation of the Size Principle—that people’s tendency to limit their generalization to the lowest level category that is consistent with the presented evidence increases with the number of exemplars observed. Counter to the Size Principle, we found that people tended to generalize a novel property displayed by a single positive observation more broadly when it was contrasted with a greater number of (and more diverse) negative exemplars than when it was contrasted with a single negative exemplar. This was found even though a narrower generalization was consistent with the observed data. Importantly, this result suggests that principles that apply to induction from positive evidence may not apply, or may need to be amended, to account for patterns in induction from mixed data that includes both positive and negative evidence. We return to this point in the General Discussion.

In this experiment, we also found that people were more likely to identify the target exemplar itself at a higher level when it was contrasted with a diverse set of negative exemplars than when it was contrasted with a single negative exemplar. We take this measure to be an indication of how participants mentally represent the target object. To be sure, the majority of participants in both conditions identified the target at the basic level (63% in the diverse condition and 75% in the nondiverse condition). Similarly, in the open-ended responses, participants overwhelmingly stated that the novel property was characteristic of a subordinate category (81% in the diverse condition and 99% in the nondiverse condition). We attribute this disproportionate responding to people’s

[2] In this and all of the experiments reported responses were only coded if they could be clearly understood as identifying the superordinate or subordinate category. Thus, the percentage of responses that identify the superordinate category (which is reported) is complementary to the percentage of responses that identify the subordinate category (which is not reported). For example, if 19% of people were coded as responding with a superordinate category, that means that 81% of people were coded as responding with a subordinate category.

[3] In this experiment and throughout, changes in degrees of freedom and in the number of people the test is based on occur when some participants failed to provide answers for the measure being tested.

prepotent tendency to identify and categorize objects at the basic level (see Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Nevertheless, the findings reported here are intriguing as they suggest contextual cues that might alter people's prepotent tendency to represent and categorize objects at the basic level (see also Lin, Murphy, & Shoben, 1997; Rosch et al., 1976).

Experiment 2

In the second experiment, we sought to conceptually replicate and extend Experiment 1's findings by changing the valence of the relevant properties, changing the categories involved, and including a behavioral measure. In this study, the categories used were fruits and vegetables which represent contrasting categories under the common parent of food. The general design was identical to that of Experiment 1 with the exceptions that we changed the categories that items were drawn from, and that the single target exemplar in this experiment cost participants points and the negative exemplars awarded points. By making the target object costly and the negative evidence rewarding, we aimed to rule out any explanation that involved valence of the property in question. We also included a new behavioral measure that asked participants to play the game for themselves and choose whether to accept or reject various foods. This measure tested whether extending the property associated with the target to members of a higher level category when shown a more diverse set of negative evidence would be reflected in people's actual behavior—namely, whether they would avoid objects from the same superordinate category as the costly target exemplar.

Method

Participants. One hundred eighty-five participants (70 female; M_{age} = 34.79, SD = 12.70) were recruited online using Amazon's Mechanical Turk. This experiment was attached to the end of an unrelated survey. The sample size for this survey was determined by the sample size of the survey to which this experiment was attached. Participants were randomly assigned to either the no diversity condition or the diverse condition.

Materials. Color pictures of six different objects all from the general category of food were used. The six objects were distributed between two contrasting categories: fruits (an apple, a banana, and a strawberry) and vegetables (carrot, broccoli, and lettuce). Depending on condition assignment, participants either saw one object from one category repeated three times (no diversity), or three different objects from one category (diverse) as the negative evidence. All participants saw one object from the other category as the target.

Procedure. The procedure for this experiment was similar to Experiment 1. As in the previous study, participants were told that they would be playing a simple game whose objective was to win as many points as possible. The instructions and the set-up of the learning phase were the same as in Experiment 1 except that in this experiment, three trials showed pictures of objects that awarded five points and one trial showed a picture of an object that cost 10 points. The order of trials was randomized across participants. Similar to Experiment 1, the distinction between the costly object and the rewarding objects established a target object and contrasting negative evidence. However, in this experiment, the costly object was the single target object while the rewarding objects were the negative exemplars. As before, the single target example was from one of the categories (e.g., fruit) while the negative examples were from the other (e.g., vegetables). For half of the participants, the single target example was a carrot and the negative examples were all fruits. For the other half of participants, the target example was an apple and the negative examples were all vegetables.

As shown in Figure 2, in the no-diversity condition the three rewarding trials each displayed the same picture of the same object (e.g., the same carrot presented three times), whereas in the diverse condition, the three rewarding trials each displayed a different object from the same contrasting category (e.g., a carrot, broccoli, and lettuce). As before, the target example (the costly object) was a single item (e.g., an apple) drawn from the category not represented by the contrasting exemplars.

Following the learning phase participants proceeded onto the test phase where they were given measures modified from Experiment 1 to be in line with the payouts described above. In the test phase, participants reported what they thought took away points in the game and chose the description that best captured how they represented the costly stimulus. We included in this experiment an additional behavioral measure: Participants were told that they would see six items selected at random and that for each item they

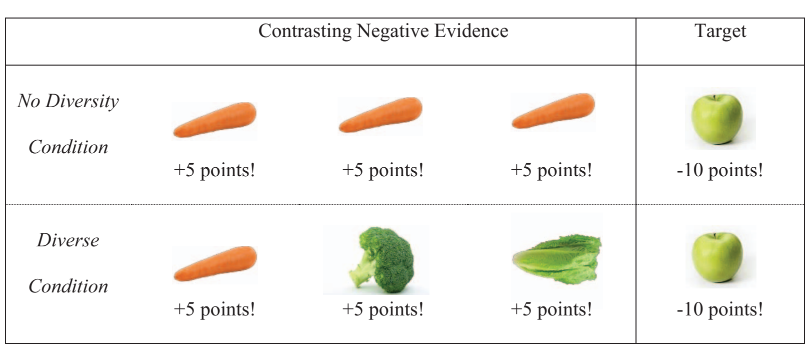

Figure 2. Depiction of the stimuli used for each condition during the learning phase of Experiment 2. The stimuli depicted above were used for participants for whom the target was an apple (the other half of the participants saw a carrot as the target and apples/fruits as the contrasts). See the online article for the color version of this figure.

would have to choose whether to select that item or pass on it. They were instructed that if they selected the item they would receive the points associated with the item (rewarding or costly), and if they passed on the item they would not. They were told that they would not receive any feedback because we wanted to see how well they could perform based on what they had learned. Participants then saw pictures of foods from each of the two contrasting categories (fruits: apple, banana, strawberry; vegetables: carrot, broccoli, and lettuce) and had to indicate whether they wanted to select that item or pass on it. Our main dependent measure was the proportion of participants in each condition who chose to pass only on foods that were members of the same superordinate category (fruits or vegetables) as the target. Finally, participants were asked to rate how likely various fruits and vegetables (apple, banana, strawberry, pineapple, carrot, broccoli, lettuce, and eggplant) were to cost them points (1 – *extremely unlikely*; 6 – *extremely likely*).

Results

We coded the open-ended question in the same manner as in Experiment 1. Again, we did not find a significant difference in the percentage of participants providing a codable response between condition (88% in the diverse condition vs. 89% in the no diversity condition, $\chi^2(1, N = 185) = .10, p = .76$). Replicating Experiment 1, we again found that a greater proportion of participants stated that a more superordinate category (fruits or vegetables) was what was costly in the diverse condition (18%) than in the no-diversity condition (6%), $\chi^2(1, N = 164) = 6.27, p < .001$. Similarly, a greater proportion of participants in the diverse condition selected the higher level description of the costly stimulus (28%) than in the no-diversity condition (9%), $\chi^2(1, N = 182) = 10.70, p < .001$. Furthermore, and new to this study, analysis of the behavioral measure showed that a greater proportion of participants in the diverse condition avoided only (and all) foods that were members of the same superordinate category as the target object (84.6%) than in the no-diversity condition (63%), $\chi^2(1, N = 183) = 11.00, p < .001$.

Finally, we averaged participants' ratings of the likelihood that an item from the target's superordinate category would cost them points ($\alpha = .91$). As predicted, participants in the diverse condition thought that other items from the same superordinate category were more likely to be costly ($M = 4.86$, $SD = 1.07$) than participants in the no-diversity condition ($M = 4.37$, $SD = 1.31$), $t(181) = 2.75, p < .01, d = .41$.

Discussion

The results of Experiment 2 conceptually replicate and extend those of Experiment 1 in two important ways. First, we demonstrated the contrast diversity effect using a different set of contrasting categories (fruits and vegetables). Second, we showed that, regardless of the valence of the property, introducing diversity into a set of negative evidence drawn from a contrasting category increases the breadth of generalization of that property from a target exemplar. We again found that more people inferred that the property associated with the target object was characteristic of a higher level category, identified the target object at a higher level, and believed that other members of the target's superordinate category were likely to possess the same property as the target when it was presented with a diverse set of contrasting negative exemplars than when it was presented with a single contrasting negative exemplar. Furthermore, and new to this experiment, a behavioral measure revealed that the diversity of negative evidence also impacted people's behavioral action tendencies (i.e., avoidance) to novel items related to the target exemplar.

Experiment 3

It might be argued that the contrast diversity effect derives from viewing negative evidence that suggest a *class* of objects versus viewing the same item repeated three times. Similarly, another alternative explanation of our results may contend that the increased number of negative exemplars in the diverse condition, rather than the degree of diversity in the negative exemplars, drove the increase in generalization. To rule out these alternative explanations, in this experiment, we replaced the no-diversity contrast set with a contrast set consisting of three different exemplars from the same basic level category (e.g., three *different* carrots). With this change in the manipulation of diversity, Experiment 3 aimed to go beyond the mere presence of diversity and demonstrate that the *degree* of diversity in the contrasting negative evidence influences the breadth of generalization from a target exemplar.

In this experiment, we also sought to expand on the previous studies by demonstrating the contrast diversity effect using a natural property—vitamins found in different foods. The strength of using points as the to-be-generalized property in Experiments 1 and 2 is that it is a truly blank property—participants had few, or no, prior expectations about the distribution of points across objects before beginning the game. However, points are clearly also an artificial property as they must have been assigned by a person (the experimenter). Thus, in the current experiment, we sought to demonstrate the contrast diversity effect with natural properties presumed to be found in the external world. Hence, we replaced the number of points each object was worth with the type of vitamin ("Precigen" or "Thiacin") each object was said to contain. Importantly, the vitamins we used were fictitious to ensure that participants did not have prior beliefs about them.

Finally, we made some minor methodological changes in this experiment to ensure that the contrast diversity effect holds across subtle variations in measurement and presentation of observations. First, we fully counterbalanced which items were presented as positive and negative evidence. This was done to ensure that the contrast diversity effect observed in Experiments 1 and 2 was not an artifact of specific characteristics of the chosen target exemplar (e.g., its typicality). Next, we replaced the game and the likelihood measure from Experiment 2 with a single measure of generalization where we simply asked participants to indicate whether or not a series of novel objects contained the property in question ("Precigin"). This change simplified the experiment for participants and provided yet another way to measure whether increasing the diversity of negative exemplars leads to increased generalization from a single positive observation.

Overall, the general set-up for the experiment was similar to the previous two except that participants saw a set of negative exemplars that was populated by a variety of items (and hence suggestive of a class), and only the degree of diversity within the negative evidence was manipulated between participants.

Method

Participants. Two hundred three participants (84 female; M_{age} = 35.66, SD_{age} = 9.81) were recruited online using Amazon's Mechanical Turk. The sample size for this survey was determined to achieve 80% power to detect a small to medium effect size (d = .40). This effect size was a conservative estimate based on the previous studies. Participants were randomly assigned to either the low-diversity condition or the high-diversity condition.

Materials. The materials used in this experiment expanded on the set of pictures used in Experiment 2. In total, color pictures of 18 different objects all from the general category of food were used. The 18 objects were distributed between two contrasting categories—fruits and vegetables—and, within those, between three subcategories. Within fruits there were three apples, three bananas, and three strawberries; and within vegetables there were three carrots, three heads of broccoli, and three cobs of corn. Depending on condition assignment, participants saw as the negative evidence either three objects from one subcategory (e.g., three apples) or three objects from different subcategories but all under the same superordinate category (e.g., one apple, one banana, and one strawberry). All participants saw one object from the other superordinate category as the target.

As in the previous experiments, which superordinate category (fruits or vegetables) the negative evidence and the target object were drawn from was counterbalanced across participants. Additionally, and new to this experiment, which specific items were shown as negative evidence and the target were fully counterbalanced across all participants (i.e., for participants who had a fruit target, some saw an apple whereas others saw a banana and still others saw a strawberry).

Procedure. The procedure was similar to the previous studies but was adapted to have vitamins as the properties in question. Participants were told that they would be playing a game where their goal was to use information that we presented to them to learn which kinds of items contain which vitamins. As in previous studies, there were two rounds: a learning phase and a test phase. Before they began, we told participants that scientists had recently discovered a new vitamin called "Precigen." They were told that in the learning phase, we would show them a sample of objects one at a time and tell them whether that object contained Precigen or a different vitamin called "Thiacin." Thus, participants were instructed that the presence of one vitamin was mutually exclusive with the presence of the other. During the learning phase, participants saw four trials. Three of the trials displayed an object with the text "Contains Thiacin" printed below it. One trial displayed an object with the text "Contains Precigen" printed below it. The Thiacin trials served as the negative evidence while the single Precigen trial was the target observation. As before, the order of trials was randomized across participants.

As shown in Figure 3, in the low-diversity condition the three Thiacin trials each displayed three different objects all from the same subordinate category (e.g., three different carrots), whereas in the high-diversity condition, the three Thiacin trials each displayed a different object from the same superordinate category (e.g., a carrot, broccoli, and corn). As before, the target example (the Precigen trial) was a single item drawn from the superordinate category not represented by the contrasting exemplars. Within this, the specific items shown were fully counterbalanced across participants.

We had two dependent measures in this study. The first was a free response question where we asked participants, "Based on what you learned, what do you think contains Precigen?" Next, participants saw a list of eight foods—apples, bananas, strawberries, oranges, carrots, broccoli, corn, celery (order of presentation was randomized)—and were asked to indicate on a binary measure whether they thought the item contained Precigen or did not. With this measure we created an indicator of whether participants extended the property "contains Precigen" to all (and only) members of the target object's superordinate category (e.g., if the target was

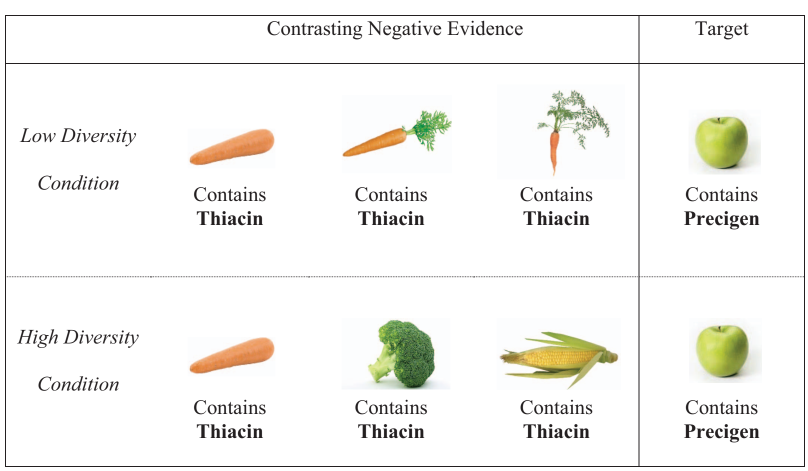

Figure 3. Depiction of the stimuli used for each condition during the learning phase of Experiment 3. These were the stimuli used for participants for whom the target was an apple. Half the participants saw a fruit as the target and vegetables as the contrasts the other half saw a vegetable as the target and fruits as the contrasts (which specific items were shown was counterbalanced across participants). See the online article for the color version of this figure.

an apple whether they also said bananas, oranges, and strawberries contained Precigen).

Results

We again did not find significant differences between conditions in the percentage of participants providing codable responses to the open ended question (89% in the high-diversity condition vs. 93% in the low-diversity condition, $\chi^2(1, N = 203) = 1.30$, $p = .25$). In analyzing the coded responses, we found that a greater percentage of participants stated that a higher level category (fruits or vegetables) was what contained Precigen in the high-diversity condition (53%) than in the low-diversity condition (18%), $\chi^2(1, N = 185) = 24.25$, $p < .001$. This result was mirrored in participants' extension of the property to other members of the target's superordinate category as indicated by which items they said contained Precigen. As predicted, we found that a greater percentage of participants generalized the property Precigen to all (and only) members of the target's superordinate category in the high-diversity condition (66%) than in the low-diversity condition (40%), $\chi^2(1, N = 203) = 14.09$, $p < .001$.

Discussion

The results of Experiment 3 support the hypothesis that the degree of diversity within a contrasting set of negative evidence influences the breadth of generalization from a target exemplar and the category level that the target is inferred to represent. For example, participants were more likely to extend the property assigned to a single apple to other fruits when it was contrasted with various vegetables than when it was contrasted with three different carrots. This was found both in people's inferences about the category of objects that contained the property ("Precigen") as well as in their generalization to specific objects within the same superordinate category as the target. Taken together, these findings go beyond numerosity and the mere presence of diversity, to suggest that greater diversity of negative evidence increases breadth of generalization from a target object while simultaneously raising the target's inferred category level.

This experiment extends the previous experiments in two additional important ways. First, it demonstrates the contrast diversity effect using a naturally occurring property (vitamins found in foods). Second, by fully counterbalancing which specific items were presented as negative and positive evidence, this experiment rules out characteristics of specific items (e.g., typicality) as the driving factor of the observed contrast diversity effect in the previous two experiments.

Taken together, Experiments 1–3 demonstrated the contrast diversity effect across a variety of measures and across variations in the presentation of these measures, which suggests that the effect is robust to subtle changes in measurement. Moreover, these experiments show that the contrast diversity effect can occur with different types of categories (sports equipment and food) and with both artificial properties (e.g., points) and naturally occurring properties (e.g., vitamins). Future research could explore how different properties and prior beliefs about those properties may interact with the diversity of negative evidence to influence induction and generalization from those observations.

Experiment 4

As we argued in the Introduction, the contrast diversity effect occurs when people rely on previously held categories that organize the negative and positive evidence into contrasting categories that share a common parent. In Experiment 4, we sought to test this theorized boundary condition of the contrast diversity effect, stating that the influence of diversity of negative evidence is constrained by the availability of a superordinate category that both unites the negative exemplars and distinguishes them from the target.

We predicted that the contrast diversity effect would not hold when the contrast set was so diverse as to suggest a higher level category that included the target exemplar. This would happen when the negative evidence was drawn from multiple contrasting categories that share a common parent with a target as opposed to a single contrasting category as was the case in the previous three experiments. For example, if a contrast set comprises a carrot, a block of cheese, and a salmon fillet, then the lowest level category that would encompass these items may be the category of food. If, in this case, the target exemplar is an apple, the contrast category of food includes the target and thus does not distinguish the contrasting exemplars from the target exemplar. In such cases, we argue, people will represent the contrasting and target items individually (or at the basic level) rather than as members of higher level, mutually exclusive, categories. Thus, we hypothesized that people would generalize more from a target exemplar when it was contrasted with a set of negative exemplars that formed a single coherent but distinct superordinate category than when the negative exemplars were drawn from multiple contrasting categories and were thus so diverse as to suggest a higher level category that encompasses both the negative exemplars and the target exemplar.

Method

Participants. Two hundred one participants (74 female; $M_{age} = 34.00$, $SD = 9.41$) were recruited online using Amazon's Mechanical Turk. The sample size in this survey was determined to achieve 80% power to detect a small to medium effect size ($d = .40$). Participants were randomly assigned to either the distinct or the nondistinct condition.

Materials. The materials consisted of eight pictures of objects all from the general category of food, drawn from the contrasting subcategories fruit, vegetables, dairy, grains, and meats. The eight pictures were the following: an apple, a carrot, broccoli, lettuce, celery, cheese, spaghetti, and a salmon fillet. Each category was represented by one of the pictures, with the exception of the vegetables category which was represented by four of the pictures.

Procedure. Having found that the contrast diversity effect holds across different categories and different types of properties (natural and artificial), we decided to return to the version of our paradigm where participants saw food objects worth different amounts of points. The procedure was similar to that of Experiment 1 with the exception of the specific stimuli used (here it was food instead of sports equipment). The learning phase in the current experiment consisted of five trials, four of which displayed pictures of objects that were worth zero points and one of which displayed a picture of an object that was worth 10 points. Again, this distinction established a target object (the rewarding object) and negative evidence (nonrewarding objects).

Because there was no effect of which superordinate category (fruits vs. vegetables) the target was drawn from in the previous experiments, we simplified the design so that all participants saw the picture of the apple for the target object. Participants in the distinct condition saw four items all from the same contrasting category of vegetables (carrot, broccoli, lettuce, and celery) presented as negative evidence (top row of Figure 4). Those in the nondistinct condition saw four food items from different contrasting categories (carrot, cheese, spaghetti, and salmon) presented as negative evidence (bottom row of Figure 4).

After the learning phase, all participants completed the same measures as in Experiment 1. Specifically, they were asked to report what they thought gives them points in the game, with responses coded based on whether they believed that the superordinate category of fruits defined the class of objects that were rewarding. Participants also rated on a 6-point scale how likely various foods were to give them points. Of interest, participants were asked how likely they thought a banana, a strawberry, and a pineapple were to give them points in the game. Finally, participants were again asked to indicate whether they represented the rewarding stimulus at a lower level as an apple or at a higher level as fruit.

Results

As before, there was not a significant difference between conditions in the percentage of participants providing a codable response to this measure (88% in the distinct condition vs. 81% in the nondistinct condition, $\chi^2(1, N = 201) = 1.79$, $p = .18$). First, coding participants' responses to the open ended question using the same method as in the previous experiments, we found that a greater proportion of participants in the distinct condition stated that the more superordinate category of fruits was what was rewarding (40%) than in the nondistinct condition (10%), $\chi^2(1, N = 170) = 20.24$, $p < .001$.

Next, we averaged participants' ratings of the likelihood that another member of the target's superordinate category would earn them points (i.e., the average rating of the likelihood that a strawberry, a banana, and a pineapple would give them 10 points; $\alpha = .94$). As predicted, participants in the distinct condition thought that other items from the same superordinate category were more likely to be rewarding ($M = 4.35$, $SD = 1.52$) than participants in the nondistinct condition ($M = 3.75$, $SD = 1.73$), $t(199) = 2.63$, $p < .01$, $d = .37$. Finally, we found that a greater proportion of participants preferred the higher level description of the target in the distinct condition (41%) than in the nondistinct condition (24%), $\chi^2(1, N = 200) = 6.59$, $p < .01$.

Discussion

In Experiment 4, we found that participants generalized a novel property from a single target (apple) less when it was contrasted with food objects from various contrasting categories (e.g., dairy, grains, meats) than when it was contrasted with various items from the single contrast category of vegetables. The results of this experiment demonstrate the theorized boundary condition wherein increased diversity within a set of negative exemplars only increases generalization from a target object when the negative observations can be covered by a single contrasting category that unites them and, critically, distinguishes them from the target. Moreover, the finding that increasing the diversity of negative evidence past a certain point can decrease generalization demonstrates a nonmonotonic relationship between diversity in negative evidence and generalization. This result is of note because it suggests another way in which induction from mixed observations (both positive and negative) does not conform to empirically established principles of induction from positive evidence. Overall, this finding supports the notion that the contrast diversity effect relies on category-based reasoning, as it suggests that the effect will only occur when people sort the negative and positive evidence into distinct categories that are represented as contrasts underneath a common superordinate category.

Experiment 5

Experiment 5 sought to demonstrate a second boundary condition to the contrast diversity effect by showing that the diversity of negative evidence will not impact generalization when the negative exemplars are drawn from a category that does not share an obvious parent with the target. Recall that in the previous four experiments, all of the items fell within the same general parent category (sports equipment in Experiment 1 and food in Experiments 2–3). In the current experiment, we varied across participants whether the target exemplar and the contrast exemplars shared a parent category (i.e., were all foods) or did not (i.e., the target was a food and the contrasts were tools). We predicted that

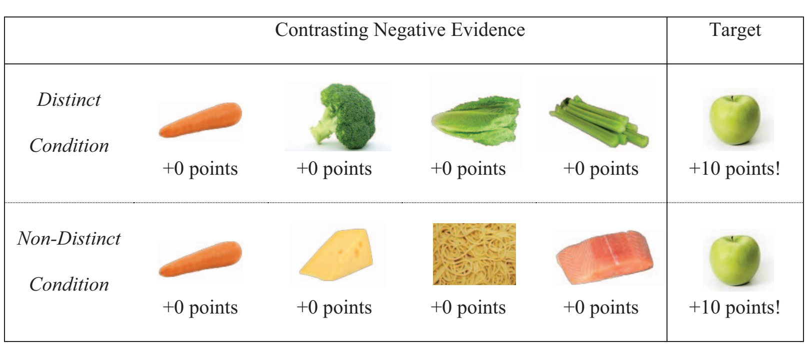

Figure 4. Depiction of the stimuli used for each condition during the learning phase of Experiment 4. See the online article for the color version of this figure.

the contrast diversity effect would not hold when the negative observations and the target observation are drawn from categories that do NOT share an obvious common parent category.

This boundary condition follows from the assumption, established in previous work (Medin et al., 2003), that the primary function of negative evidence is to highlight the relevant category level for induction. Thus, the degree of diversity in negative evidence that is drawn from a closely related, but mutually exclusive subcategory (i.e., a single contrast category) should have a larger impact on inductive reasoning since its close relation to, but exclusion of, the target should do the most to make a given category level salient (see also Gentner & Markman, 1994).

Method

Participants. Two hundred nineteen participants (141 female; M_{age} = 36.77; *SD* = 12.60) were recruited online using Amazon's Mechanical Turk. The experiment was administered at the end of an unrelated survey. As a result, the sample size of this study was determined by the sample size of the study to which it was attached. This experiment employed a 2 (diversity of negative evidence: low vs. high) × 2 (parent category of negative evidence: same as target vs. different than target) between subjects design. Participants were randomly assigned to condition.

Materials. The materials consisted of 11 pictures in total. A picture of an apple served as the target object in all conditions. Of the 10 remaining pictures, five shared the parent category of food (all vegetables: three different carrots, broccoli, and lettuce) with the target and five did not (all tools: three different hammers, a screwdriver, and a wrench). These materials allowed us to create four different contrast sets as depicted in Figure 5.

Procedure. The procedure used was similar to that of Experiment 3. In this experiment, the three contrasting negative exemplars *cost* five points and the single target exemplar *rewarded* 10 points. Because of an oversight, order of trials was not randomized in this experiment and the target exemplar always appeared in the fourth position.

To manipulate diversity in the low diversity condition, the negative exemplars were three different items all from the same basic level category (carrots or hammers). In the high diversity condition, the negative exemplars were three different objects all from the same superordinate category (vegetables or tools). In both conditions, the target was an apple.

Whether the contrasting negative observations were drawn from the same parent category or a different parent category was orthogonally manipulated between participants as well. The parent category of the target exemplar was food. So, participants in the same parent category condition saw contrasts that were all vegetables. Those in the different parent category condition saw contrasts that were all examples of tools. Figure 5 illustrates the four conditions of the crossed design.

During the test phase, participants answered the open-ended question of what gave them points in the game. Next, participants were asked which level of description best described the rewarding object. The descriptions ranged from the basic level (apple) to the superordinate (fruit) to the parent of that superordinate (food). Finally, participants rated how likely various items (apple, orange, banana, potato, celery, wrench, saw, pliers) were to give them points.

Results

To analyze participants' responses to the open-ended question of what they thought gave them points, we coded their responses based on whether they inferred that the target represented a category above the basic level. Participants who wrote that either the superordinate (i.e., fruits) or the parent category (i.e., food) was the rewarding category received a code of 1; participants who said that

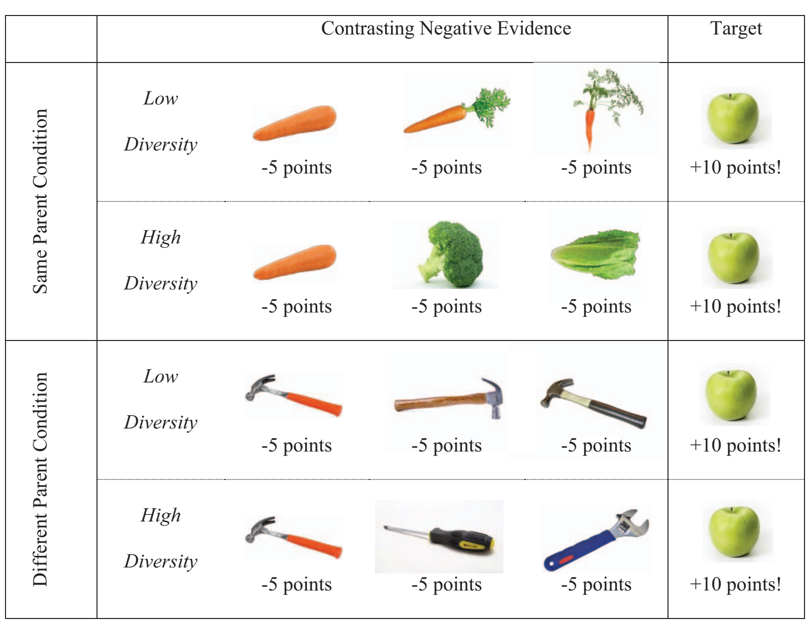

Figure 5. Depiction of the stimuli used for each condition during the learning phase of Experiment 5. See the online article for the color version of this figure.

it was the basic level category (i.e., apple) received a code of 0.[4] As before, we coded more subordinate categories (e.g., green apple) as 0 and responses that listed two categories (e.g., apples/fruit) based on the highest level category listed. Again, anything other than these responses was not coded and left as missing (n = 25, 11% of responses). Neither diversity condition, parent condition, nor their interaction had a significant impact on the likelihood of participants providing a codable response (all ps $>$ .15).

A binary logistic regression analysis revealed a main effect of diversity (b = .59, SE = .20, Wald χ^2 = 8.75, $p < .01$) such that increased diversity in contrasts increased the likelihood that participants inferred that the target corresponded to a higher level category (i.e., above the basic level). This result replicates our previous experiments. We did not find a main effect of parent condition (b = −.01, SE = .20, Wald χ^2 = .01, p = .96). Most importantly, we found that the main effect of diversity was qualified by a significant interaction, b = −.40, SE = .20, Wald's χ^2 = 4.12, p = .04 (see Figure 6). Breaking this interaction down and looking at the simple effects, we found that increasing the diversity of the contrasts increased the likelihood that participants inferred that the target represented a higher level category when the contrasts and target shared a common parent (b = .99, SE = .30, Wald χ^2 = 10.77, p = .001), but not when the contrasts and targets came from different parent categories (b = .18, SE = .26, Wald χ^2 = .51, p = .48).

Next, we performed the same analysis on the forced choice question that asked participants to choose which identification of the rewarding stimulus they preferred. We coded participants' choices in the same way as in previous experiments to create a variable that indicated whether participants identified the target at the basic level (i.e., apple) or at any level above that (i.e., fruit or food). Using a binary logistic regression, we found only a main effect of diversity (b = .41, SE = .16, Wald χ^2 = 6.44, p = .01) such that people were more likely to identify the stimulus at a higher level when the contrasts were more diverse. Neither the main effect of parent condition (b = −.25, SE = .16, Wald χ^2 = 2.51, p = .11) nor the interaction (b = .01, SE = .16, Wald χ^2 = .01, p = .94) were significant.

Finally, to explore the effects of contrast diversity and parent category on generalization, we averaged participants' ratings of the likelihood that items from the target's superordinate category would earn them points (i.e., the average likelihood that an orange and a banana would be rewarding; these items were highly consistent with each other α = .94). We conducted a two-way ANOVA and found a main effect of diversity, $F(1, 215) = 14.62$, $p < .001$, $\eta^2_p = .06$, replicating our previous findings; a marginal main effect of parent condition $F(1, 215) = 2.84$, $p = .09$, η^2_p = .01; and importantly, a significant interaction $F(1, 215) = 4.01$, $p = .05$, $\eta^{2-}{}_{p} = .02$ (see Figure 7). Breaking down this interaction, we found a significant simple effect of diversity when the negative evidence was from the same parent category as the target, $F(1, 215) = 17.21$, $p < .001$, $\eta^2_p = .07$, such that participants rated other members of the target's superordinate category as more likely to be rewarding in the high diversity condition than in the low diversity condition. We did not find an effect of diversity when the negative evidence came from a different parent category than the target, $F(1, 215) = 1.64$, $p = .20$, $\eta^2_p < .01$.

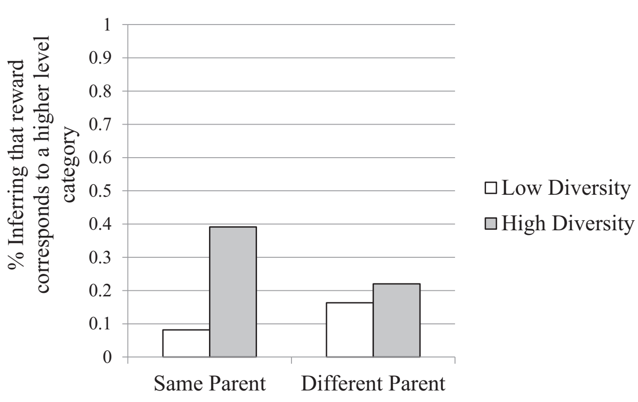

Figure 6. Percent of participants within each condition who reported that a higher level category (i.e., above basic level) was what gave rewards in the game in Experiment 5.

Discussion

This experiment set out to explore a second boundary condition to our proposed contrast diversity effect. Specifically, we examined whether increased diversity of negative evidence would impact generalization from a target exemplar when the contrasting negative exemplars and the target did not share a common parent category. As predicted, we replicated our previous finding that increased diversity in negative evidence leads to greater generalization from a target when the negative evidence and target shared a common parent category, but found that the contrast diversity effect did not occur when the contrasts and target were drawn from different parent categories. Although we did not observe this boundary condition on the measure that asked people which identification they preferred, we take the overall results of this experiment as supporting the hypothesis that the degree of diversity within a contrast set does not have an—or at least has an attenuated—effect on generalization when the target and contrasts come from superordinate categories that do not share a common parent. Overall, the boundary conditions observed here and in Experiment 4 support the notion that hierarchical representations of category structures are critical for producing the contrast diversity effect.

General Discussion

In this research, we established and explored the contrast diversity effect—the effect of diversity of negative evidence on the inductive inferences drawn from a single observation of a target exemplar. In Experiments 1 through 3, we established the contrast

[4] The data were coded in this way to create a variable that indicated whether people inferred the target represented the basic level category or anything higher. Note that the highest level option *food* is not equally relevant to the same parent and different parent condition. For participants in the same parent condition, the category *food* is ruled out by the contrast exemplars (which are vegetables) that are also from the parent category food. Thus, saying that the category *food* is rewarding would not be a reasonable answer for these participants. (As an additional note: only one participant in the same parent condition said "food" and the results reported above hold when excluding this participant.) This asymmetry in plausible answers led us to code responses to indicate any movement away from the basic level category as this measure is not sensitive to these potential differences between conditions.

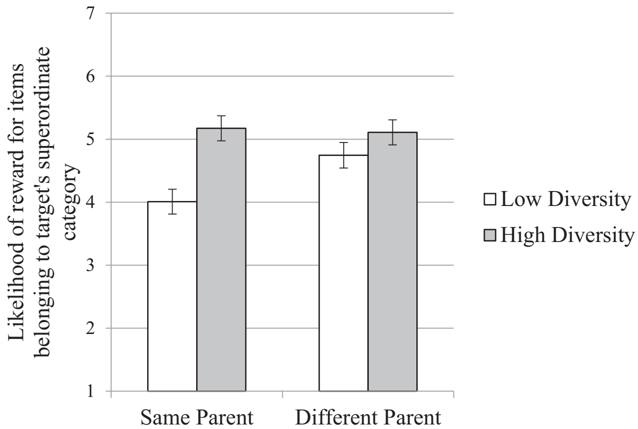

Figure 7. Mean rating of the likelihood that other members of the target's superordinate category would be rewarding for Experiment 5. Higher numbers indicate a higher likelihood rating and thus greater generalization of the reward property of the target. Error bars represent one standard error above and below the mean.

diversity effect by showing that increasing the diversity of negative evidence drawn from a contrasting category led people to infer that a target exemplar corresponded to a higher level category and led to greater generalization of a novel property associated with the target. Importantly, we demonstrated the contrast diversity effect using sets drawn from different parent categories (sports equipment in Experiment 1, and food in Experiments 2 and 3). In Experiment 2, we further showed that the effect does not depend on the valence of the property in question. In Experiment 3, we showed that even when holding the number of objects presented as negative evidence constant, greater diversity among them leads to greater generalization from a target object. Furthermore, whereas Experiments 1 and 2 used points as a novel and blank property to be induced, Experiment 3 extended these results by demonstrating the contrast diversity effect using a naturally occurring property—vitamins. These results suggest that the contrast diversity effect depends on the degree of diversity of negative evidence, and is not unique to a single set of categories or a single type of property.

In Experiments 4 and 5, we tested two boundary conditions to the contrast diversity effect that follow from a category-based account of inductive reasoning. Experiment 4 showed that the contrast diversity effect only occurred when the negative evidence was consistent with a single category that both united the negative exemplars and distinguished them from the target exemplar. Finally, in Experiment 5, we found that the contrast diversity effect was limited to instances in which the negative evidence and the target were drawn from contrasting categories that shared an obvious parent category. Overall, these findings demonstrate that increasing the diversity of negative evidence alone increases generalization from a target so long as the negative evidence is drawn from a single contrast category that excludes, but shares a common parent with, the target.

Blank Versus Nonblank Properties

In the current research, we deliberately tested the contrast diversity effect using blank properties. By using points and fictional vitamins across all of our studies at the properties in question, we sought to avoid complications that might arise from participants holding prior beliefs about the distributions of the properties. This allowed us to focus on how people use the objects and their corresponding category membership as guides for inductive reasoning.

Within the blank properties we used in this research, there is also a distinction to be made between artificial properties (points) and natural properties (vitamins). As an artificial property, points are clearly assigned to objects by an agent—in this case, the experimenter. This in itself may lead observers to make certain inferences regarding what is intended to be revealed by these specific observations (see the Sampling Assumptions section below). Although artificial, we argue that this type of induction is not uncommon. People assign arbitrary properties to objects all the time (e.g., the value of different types of paper money, the meaning of certain symbols, etc.), and we would expect the contrast diversity effect to generalize to induction in such domains. Moreover, we moved beyond artificial properties and demonstrated the contrast diversity effect with a natural property (vitamins) in Experiment 3. In line with classic induction paradigms (e.g., Osherson et al., 1990), the vitamins still constituted blank properties as they were made up and thus participants could not bring previous knowledge about the property to bear on the induction task. Future research could explore how prior knowledge/beliefs about the properties in question interacts with diversity in negative evidence to influence induction and generalization.

Sampling Assumptions

Past literature has shown that people's inductive inferences depend heavily on their assumptions of how the examples they observed were generated (see also Xu & Tenenbaum, 2007a). In the research we presented here, we designed the experiments to convey strong sampling. That is, we assume that participants inferred that the samples they observed were generated by a knowledgeable and intentional source (e.g., Grice, 1975; Medin et al., 2003; Voorspoels et al., 2015; Xu & Tenenbaum, 2007a). In the case of the Experiments 1, 2, 4, and 5, this is most strongly conveyed by the fact that the property in question was number of points which participants were likely to assume had been intentionally assigned by the experimenter. However, given the observations shown (e.g., three negative observations of the same carrot in Experiment 2), we cannot assume that participants inferred a maximally helpful teacher. One could reasonably argue that a helpful teacher would not use three out of four chances for observation on negative exemplars. Nonetheless, the instructions and context for the study suggest that the observations shown were selected by a knowledgeable source with the deliberate intention to reveal a pattern through the observations selected. Although it is beyond the scope of the current investigation, it would be an interesting avenue for future research to explore how participants' sampling assumptions influence the extent to which the diversity of negative evidence informs their inductive inferences from a single target.

Violating the "Size Principle" and the "Monotonicity Principle"

Throughout the experiments reported in this paper, we documented a consistent violation of the Size Principle which states that more narrow extensions (i.e., less generalization) are more likely than broader extensions (i.e., greater generalization) when both are consistent with a given set of examples and, further, that the preference for narrower generalizations increases exponentially with the number of examples observed (see Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007a). In our studies, we repeatedly found that increasing the diversity of contrasting negative exemplars increased generalization from a target exemplar to a broader category—even when a narrower category was consistent with the data. In Experiments 1 and 2, we found this effect when the diverse set included more unique objects than the nondiverse set, which directly contradicts the second clause in the size principle that increasing the number of exemplars seen increases the preference for more narrow generalization. Furthermore, while previous research has shown that the size principle does not hold under all sampling conditions (see Xu & Tenenbaum, 2007a), we demonstrated a violation of the size principle under strong sampling conditions—when the data was assumed to be generated by a knowledgeable source—where it has been previously shown that the size principle holds (e.g., Xu & Tenenbaum, 2007a). Thus, our results show that violations of the size principle can occur in cases where the size increase comes from an increase in the number of negative evidence exemplars drawn from a category that contrasts with a target observation.

In addition, Experiment 4 clearly demonstrates a nonmonotonic relationship between diversity in negative evidence and generalization. This finding is noteworthy because previous research focused on induction from positive evidence has consistently demonstrated that increasing the diversity of positive evidence (i.e., the breadth of positive observations) monotonically increases generalization (Osherson et al., 1990; Sloman, 1993). Our Experiments 1–3 mirror this pattern by showing that increasing the diversity of negative evidence increases generalization from a target. However, in Experiment 4, we show that increasing the diversity of negative evidence past a certain point—namely, past the point where it can be covered by a single superordinate category that excludes the target—actually decreases generalization. Taken together, this reveals a nonmonotonic relationship as increasing diversity in negative evidence does not monotonically increase generalization from a target. This suggests that the relationship between diversity in negative evidence and generalization may not follow the same function as the relationship between diversity in positive evidence and generalization (for further discussion of monotonicity in reasoning from negative evidence see Heussen et al., 2011; Kalish & Lawson, 2007; Voorspoels et al., 2015).

Our aim in highlighting these violations of previously established principles of induction is not to suggest that they are inaccurate, but rather to compare patterns in inductive reasoning from positive evidence to patterns in inductive reasoning from mixed evidence. Moreover, we do not believe that induction from negative evidence follows a process that is necessarily distinct from induction from positive evidence. Instead, we argue that the induction from positive and negative evidence follows the same process of category-based generalization, but that this process may lead to distinct patterns of generalization depending on the sign (positive or negative) of the observed evidence.

The Role of Categories in Induction From Negative Evidence

Whereas some models of induction rely heavily on the use of categories to guide and constrain induction (e.g., Coley et al., 1997; Osherson et al., 1990), others rely primarily on judgments of similarity between observed evidence and potential targets as a guide for generalization (e.g., Sloman, 1993). We believe that a model of induction that relies on similarity *alone* cannot account for our findings, and that prior knowledge about the structure of the world, in terms of category membership, appears to play an important role in inductive reasoning from negative evidence.

As Experiments 4 and 5 highlight, the contrast diversity effect depends on the accessibility of a category that complements the category implied by the negative evidence. Drawing on previous work, we argue that this is the case because the primary function of negative evidence is to highlight the relevant category level for induction (Medin et al., 2003; Voorspoels et al., 2015). When the negative evidence fails to make salient an obvious category for the target, as would be the case when it is drawn from disparate categories (Experiment 4) or noncontrasting categories (Experiment 5), the diversity of negative evidence no longer impacts induction. Overall, the observation of these boundary conditions suggests that the contrast diversity effect relies on the presence and activation of prior knowledge that organizes the negative and positive evidence within contrasting taxonomic categories.

At a theoretical level, we argue that a process of category-based induction is a more parsimonious explanation of the contrast diversity effect than would be a feature-based account that relies on calculations of similarity between negative and positive evidence. Relying on hierarchical representations of category structure saves an individual the work of recalculating feature similarity every time they encounter objects in the world. Because people hold these hierarchical representations of category structure, they can use the category level implied by the diversity in negative evidence as a cue to define the hypothesis space for generalizing from the target exemplar. That is, the inferred category level of the negative examples constrains one's search for the appropriate category level for the target by imposing the heuristic that the two will be equivalent in terms of hierarchical level. This constraint limits the likely candidates for a target category, thus making the search faster and more economical. Indeed, exploiting prior knowledge to make sense of novel situations is one of the most adaptive features of human cognition (Murphy & Medin, 1985).

Relevance Theory

Our proposal, that negative evidence serves to highlight a relevant category for induction, is consistent with the relevance theory of induction (Medin et al., 2003). The main idea behind this theory is that when reasoning from given evidence, people assume that the information provided is relevant for the problem at hand. As a result, people search for what it is within the presented information that provides a relevant basis for induction (Medin et al., 2003). This idea is especially pertinent to social situations where the evidence is presented by a knowledgeable teacher (or an experimenter). In these cases, people may assume that the teacher is following conversational norms of relevance and informativeness (cf. Cruse, 1977; Grice, 1975) and is only presenting information that is useful for the task at hand. Assuming that a teacher is attempting to be helpful, or at least is intentional in selecting observations, an individual may reasonably infer that any presented negative evidence is presented as a counter example that highlights the relevant category for induction by demarcating the boundary between categories to which the property does and does not extend (Heussen et al., 2011; Voorspoels et al., 2015). So, for example, the inclusion of various vegetables as negative evidence may serve to highlight the commonality between an apple and objects such as a banana, or a strawberry in that they belong to the same subcategory of food, that is, fruit, while simultaneously informing participants that the property does not extend to other subcategories of food such as vegetables.

Overhypotheses

The research presented here also relates to previous work detailing the inferential mechanism of *overhypothesis*, which describes hypotheses that people form about the dimensions that are characteristic of a set of categories (e.g., all animals make a characteristic sound, or different food groups have different reward values). Such overhypotheses are hypotheses that involve second-order generalizations about abstract qualities of categories (Goodman, 1955/1983; Shipley, 1993). First-order generalizations are those made about specific properties found in a given category. For example, the generalization that all vegetables are worth zero points is a first-order generalization because it is about a specific property within a single category. Second-order generalizations abstract from first order generalizations to make projections about the *types of properties* that are characteristic of a set of categories (for further discussion see Shipley, 1993). For example, abstracting from the first-order generalization that vegetables are worth zero points to infer that different food groups have characteristic reward values is a second-order generalization, and thus an overhypothesis. Overhypotheses have been shown to support inductive inferences about the prevalence of properties in novel target categories based on information about contrasting categories (Dewar & Xu, 2010; Kemp, Perfors, & Tenenbaum, 2007; Macario, Shipley, & Billman, 1990; Nisbett et al., 1983). For instance, given the overhypothesis that each food group is associated with a given reward amount, it takes only one observation of an apple winning 10 points to infer that fruits give 10-point rewards.

One way to interpret our results is that people use the diversity displayed by the exemplars provided for a contrasting category as a cue to the amount of diversity to expect in the target category. This inference, which can be thought of as an overhypothesis about the degree of diversity within each category, was reflected in the level at which people represented the target category. Specifically, we found that people used the level implied by the contrast category to infer the appropriate category level along which to extend overhypotheses about properties. Within these results, category diversity and category level can each be thought of as abstract qualities of categories that have been generalized from one category (the contrast) to another (the target). More generally, our findings suggest that overhypotheses can involve inferences about *distributions* of populations and properties in addition to inferences about properties and dimensions themselves (see also Kemp et al., 2007).

Practical Relevance

Finally, it is important to consider the practical relevance of the findings presented here. Because the primary goal of this research is to document a consistent pattern in reasoning that sheds light on the basic processes involved in induction, the tasks and observations we presented to participants occurred in a rather contrived setting. One may wonder whether and when people would face a situation where they encounter multiple negative exemplars and then are tasked with generalization from a single positive observation. As discussed earlier, it would seem unlikely that a teacher (or anyone else) would present a multiplicity of negative exemplars and only a single positive exemplar if her goal was to teach a student about the prevalence of a property displayed by a positive observation. However, it may still be relatively common that people encounter a wealth of information about the prevalence of a given property within one category and subsequently observe a novel item possessing a different property. For example, a teacher may present an observer with several examples of items possessing a given property and a single contrasting example with the intent of teaching about the more observed property; but an observer may, for various reasons, be more interested in the divergent property observed in a single negative exemplar. In this case, the observer's extent of generalization from the single item may be influenced by the diversity of items they were shown to possess the property that the teacher intended to teach him/her about.

Generally, we argue that the contrast diversity effect may be relevant in any situation that meets the following preconditions: (a) an individual holds a clear representation of some taxonomy of categories relevant to the observations at hand; (b) that individual has encountered many observations of a given property; and (c) that same individual is then presented with a novel observation of an object within that same taxonomy but that possess a divergent property than what has previously been observed. In this case, we would expect the diversity of observations of the first property to influence people's generalization of the divergent property to unencountered items.

One prominent example is social inference. Specifically, social life represents an important domain where people may use the diversity of observations possessing one property to guide inferences about the prevalence of a distinct property. For example, suppose an individual has grown up in a country with clear social norms and conventions such as shaking hands to greet people. From her experience observing her neighbors all display this behavior, that individual may induce that shaking hands is the conventional way that people greet each other. Now suppose this individual meets a new person visiting from a different country and this person greets her by kissing each cheek. The local person may subsequently induce that all people from the visitor's country greet by kissing each other on the cheek. In this case, the local has used information about the prevalence of shaking hands within her own country as negative evidence to induce the scope of the novel behavior—kissing cheeks—observed by a single encounter. To the extent that people are represented as members of groups that can be structured under superordinate categories (e.g., race, culture, nationality), one could reasonably predict that patterns of induction demonstrated in basic paradigms may also occur in social inferences about the prevalence of traits or norms across groups of individuals (see also Cooley & Payne, 2018; Riggs, 2019). As another example, the present research may be especially relevant in cases where people reason from experience with a token minority to make generalizations about an entire group of people. If the token individual displays a behavior or trait that is novel, we would predict that people's generalization of that behavior/trait would be influenced by the diversity of observations they have had of people from contrasting groups displaying a different trait/behavior. Exploring the impact of negative and contrasting evidence on stereotyping and generalization in social domains is an intriguing avenue for future research (see also Riggs, 2019).

Concluding Remarks

Taken together, we have shown that people raise the level of the category that a target object is inferred to represent and thus generalize more broadly from a target object, when it is contrasted with a more diverse set of negative evidence drawn from a mutually exclusive subcategory that shares a common parent with the target. These findings may have far-reaching implications for how people learn from their experiences. Specifically, this work suggests that not only do people learn and make predictions based on information gathered from observation of a target, but they may also make inferences about targets based solely on characteristics (in this case, diversity) of contrasting observations.

References

Billman, D., & Dávila, D. (2001). Consistent contrast aids concept learning. *Memory & Cognition, 29,* 1022–1035. http://dx.doi.org/10.3758/BF03195764

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. *Cognition, 64,* 73–112. http://dx.doi.org/10.1016/S0010-0277(97)00017-6

Cooley, E., & Payne, B. K. (2018). A group is more than the average of its parts: Why existing stereotypes are applied more to the same individuals when viewed in groups than when viewed alone. *Group Processes & Intergroup Relations*. Advance online publication. http://dx.doi.org/10.1177/1368430218756494

Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics, 13,* 153–164. http://dx.doi.org/10.1017/S0022226700005363

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge. Evidence from 9-month-old infants. *Psychological Science, 21,* 1871–1877. http://dx.doi.org/10.1177/0956797610388810

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5,* 152–158. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00652.x

Goodman, N. (1983). *Fact, fiction, and forecasts*. New York, NY: Bobbs-Merrill. (Original work published 1955)

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *WIREs Cognitive Science, 1,* 278–292. http://dx.doi.org/10.1002/wcs.44

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569–592. http://dx.doi.org/10.3758/BF03212996

Heussen, D., Voorspoels, W., Verheyen, S., Storms, G., & Hampton, J. A. (2011). Raising argument strength using negative evidence: A constraint on models of induction. *Memory & Cognition, 39,* 1496–1507. http://dx.doi.org/10.3758/s13421-011-0111-2

Kalish, C. W., & Lawson, C. A. (2007). Negative evidence and inductive generalization. *Thinking & Reasoning, 13,* 394–425. http://dx.doi.org/10.1080/13546780701273402

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10,* 307–321. http://dx.doi.org/10.1111/j.1467-7687.2007.00585.x

Lin, E. L., Murphy, G. L., & Shoben, E. J. (1997). The effects of prior processing episodes on basic-level superiority. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50,* 25–48. http://dx.doi.org/10.1080/713755686

López, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition, 23,* 374–382. http://dx.doi.org/10.3758/BF03197238

Macario, J. F., Shipley, E. F., & Billman, D. O. (1990). Induction from a single instance: Formation of a novel category. *Journal of Experimental Child Psychology, 50,* 179–199. http://dx.doi.org/10.1016/0022-0965(90)90038-A

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language, 31,* 195–217. http://dx.doi.org/10.1016/0749-596X(92)90011-L

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 54–70. http://dx.doi.org/10.1037/0278-7393.23.1.54

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10,* 517–532. http://dx.doi.org/10.3758/BF03196515

Murphy, G. L. (2010). What are categories and concepts? In D. Mareschal, P. C. Quinn, & S. Lea (Eds.), *The making of human concepts* (pp. 11–27). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199549221.003.02

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316. http://dx.doi.org/10.1037/0033-295X.92.3.289

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363. http://dx.doi.org/10.1037/0033-295X.90.4.339

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200. http://dx.doi.org/10.1037/0033-295X.97.2.185

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. http://dx.doi.org/10.1037/0278-7393.26.4.811

Riggs, A. E. (2019). Social statistics: Children use statistical reasoning to guide their inferences about the scope of social behavior. *Developmental Psychology, 55,* 66–79. http://dx.doi.org/10.1037/dev0000618

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior, 14,* 665–681. http://dx.doi.org/10.1016/S0022-5371(75)80055-7

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. http://dx.doi.org/10.1016/0010-0285(76)90013-X

Shipley, E. F. (1993). Categories, hierarchies, and induction. *Psychology of Learning and Motivation, 30,* 265–301. http://dx.doi.org/10.1016/S0079-7421(08)60299-6

Shipley, E. F., & Kuhn, I. F. (1983). A constraint on comparisons: Equally detailed alternatives. *Journal of Experimental Child Psychology, 35,* 195–222. http://dx.doi.org/10.1016/0022-0965(83)90079-6

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280. http://dx.doi.org/10.1006/cogp.1993.1006

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of novel stimuli in well-known natural concepts: A case study. *Psychonomic Bulletin & Review, 8,* 377–384. http://dx.doi.org/10.3758/BF03196176

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. http://dx.doi.org/10.1017/S0140525X01000061

Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). The role of contrast categories in natural language concepts. *Journal of Memory and Language, 44,* 618–643. http://dx.doi.org/10.1006/jmla.2000.2748

Voorspoels, W., Navarro, D. J., Perfors, A., Ransom, K., & Storms, G. (2015). How do people learn from negative evidence? Non-monotonic generalizations and sampling assumptions in inductive reasoning. *Cognitive Psychology, 81,* 1–25. http://dx.doi.org/10.1016/j.cogpsych.2015.07.001

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 5012–5015. http://dx.doi.org/10.1073/pnas.0704450105

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10,* 288–297. http://dx.doi.org/10.1111/j.1467-7687.2007.00590.x

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as Bayesian inference. *Psychological Review, 114,* 245–272. http://dx.doi.org/10.1037/0033-295X.114.2.245

Appendix

Item-Level Analyses

The tables below show item-level analyses for the generalization measure administered in each experiment. In each of the tables, rows display statistics for each item and columns are labeled to identify whether the statistics represent responses from participants in the low diversity or high diversity condition.

The statistics listed correspond to participants responses to the specific item only if it was in the same superordinate category as the single piece of positive evidence that participants saw (e.g., the "Baseball Hat" row in Table A1 only displays statistics for when the target object was a baseball bat and excludes participants who saw a hockey stick as the target). This was done because several of the experiments counterbalanced the superordinate category the target observation and the negative evidence was drawn from.

Table A1
Item-Level Analyses for Experiment 1

Item	$N_{Low\ Diversity}$	$M\ (SD)_{Low\ Diversity}$	$N_{High\ Diversity}$	$M\ (SD)_{High\ Diversity}$	*t* test	*d*	95% CI of *d*
Baseball hat	50	4.40 (1.52)	51	4.80 (1.56)	$t(99) = 1.31, p = .19$	.26	[−.14, .66]
Baseball jersey	50	4.16 (1.61)	51	4.31 (1.86)	$t(99) = .44, p = .66$	.09	[−.31, .49]
Baseball cleats	50	4.22 (1.61)	51	4.45 (1.80)	$t(99) = .68, p = .50$	.14	[−.26, .54]
Hockey helmet	50	4.18 (1.56)	49	4.92 (1.46)	$t(97) = 2.43, p = .02$	.49	[.09, .90]
Hockey jersey	51	4.35 (1.43)	49	4.92 (1.48)	$t(98) = 1.94, p = .05$	.39	[−.01, .79]
Hockey gloves	51	4.31 (1.48)	49	4.88 (1.60)	$t(98) = 1.83, p = .07$	.37	[−.03, .77]

Note. Each row presents the number, the mean, and the standard deviation of participants' ratings of how likely that specific item was to possess the target property (in this case winning points in the game). This measure was presented on a 6-point scale (1 = *extremely unlikely*; 6 = *extremely likely*). For each item, we conducted a *t* test to examine whether participants' likelihood of generalizing the novel property to that item differed depending on whether it was contrasted with nondiverse or diverse negative evidence.

(*Appendix continues*)

Table A2
Item-Level Analyses for Experiment 2

Item	$N_{Low\ Diversity}$	$M\ (SD)_{Low\ Diversity}$	$N_{High\ Diversity}$	$M\ (SD)_{High\ Diversity}$	*t* test	*d*	95% CI of *d*
Banana	44	4.41 (1.47)	46	4.74 (1.42)	$t(88) = 1.08, p = .28$	.23	[−.19, .65]
Strawberry	44	4.25 (1.48)	46	4.83 (1.31)	$t(88) = 1.96, p = .05$	.41	[−.01, .84]
Pineapple	44	4.45 (1.47)	46	4.85 (1.33)	$t(88) = 1.33, p = .19$	.28	[−.13, .71]
Broccoli	48	4.42 (1.33)	45	5.02 (.97)	$t(91) = 2.49, p = .01$	.52	[.11, .94]
Lettuce	48	4.46 (1.22)	45	4.89 (.96)	$t(91) = 1.88, p = .06$	.39	[−.02, .81]
Eggplant	48	4.25 (1.61)	45	4.84 (.93)	$t(91) = 2.17, p = .03$	.45	[.04, .87]

Note. Each row presents the number, the mean, and the standard deviation of participants' ratings of how likely that specific item was to possess the target property (in this case losing points in the game). This measure was presented on a 6-point scale (1 = *extremely unlikely*; 6 = *extremely likely*). For each item, we conducted a *t* test to examine whether participants' likelihood of generalizing the novel property to that item differed depending on whether it was contrasted with nondiverse or diverse negative evidence.

Table A3
Item-Level Analyses for Experiment 3

Item	$N_{Low\ Diversity}$	$\%Yes_{Low\ Diversity}$	$N_{High\ Diversity}$	$\%Yes_{High\ Diversity}$	χ^2 test
Apple	36	58%	31	81%	$\chi^2(1) = 3.84, p = .05$
Banana	35	51%	34	85%	$\chi^2(1) = 9.11, p < .01$
Strawberry	35	63%	33	97%	$\chi^2(1) = 12.09, p = .001$
Orange	53	60%	49	88%	$\chi^2(1) = 9.81, p < .01$
Carrot	33	42%	35	89%	$\chi^2(1) = 16.16, p < .001$
Broccoli	37	46%	31	74%	$\chi^2(1) = 5.56, p = .02$
Corn	34	41%	32	66%	$\chi^2(1) = 3.96, p = .05$
Celery	52	54%	49	76%	$\chi^2(1) = 5.16, p = .02$

Note. Each row presents the percentage of participants within each condition who responded "Yes" to the question of whether that specific item possess the target property (in this case "Precigen"). The last column presents a chi-square test of whether the percentage of participants generalizing to that specific item are significantly different across the low and high diversity conditions. Statistics in each row correspond to cases where that specific item was in the same superordinate category as the target observations but was not the target observation itself (e.g., if a participant saw an Apple as the target observation, they are excluded from all analyses presented in the "Apple" row of this table).

Table A4
Item-Level Analyses for Experiment 4

Item	$N_{Low\ Diversity}$	$M\ (SD)_{Low\ Diversity}$	$N_{High\ Diversity}$	$M\ (SD)_{High\ Diversity}$	*t* test	*d*	95% CI of *d*
Banana	42	3.64 (1.82)	39	4.90 (1.41)	$t(79) = 3.45, p < .001$	.78	[.33, 1.22]
Strawberry	42	3.86 (1.83)	40	5.00 (1.40)	$t(80) = 3.17, p < .01$	.77	[.26, 1.15]
Pineapple	42	3.81 (1.84)	40	4.85 (1.44)	$t(80) = 2.84, p < .01$	.64	[.19, 1.09]

Note. Each row presents the number, the mean, and the standard deviation of participants' ratings of how likely that specific item was to possess the target property (in this case winning points in the game). This measure was presented on a 6-point scale (1 = *extremely unlikely*; 6 = *extremely likely*). For each item, we conducted a *t* test to examine whether participants' likelihood of generalizing the novel property to that item differed depending on whether it was contrasted with negative exemplars that could be covered by a superordinate category that excluded the target observation (distinct) or did not (nondistinct).

(*Appendix continues*)

Table A5
Item-Level Analyses for Experiment 5

Item	$N_{Low\ Diversity}$	$M\ (SD)_{Low\ Diversity}$	$N_{High\ Diversity}$	$M\ (SD)_{High\ Diversity}$	t test	d	95% CI of d
Same parent condition							
Banana	55	4.20 (1.74)	55	5.18 (1.31)	$t(108) = 3.35, p < .001$	.64	[.26, 1.02]
Orange	54	3.91 (1.88)	55	5.16 (1.29)	$t(107) = 4.08, p < .001$	.77	[.41, 1.16]
Different parent condition							
Banana	53	4.68 (1.50)	55	5.09 (1.44)	$t(106) = 1.45, p = .15$	.28	[−.10, .67]
Orange	53	4.81 (1.46)	55	5.13 (1.42)	$t(106) = 1.14, p = .26$	.22	[−.16, .61]

Note. The table is split by whether participants saw negative exemplars that were in the same parent category as the target (i.e., Same Parent Condition means that both the target and the negative exemplars were food. Different Parent Condition means that the target was food and the negative exemplars were tools.). Each row presents statistics for participants' ratings of how likely that specific item was to possess the target property (in this case winning points). For each item, we conducted a *t* test to examine whether participants' likelihood of generalizing the novel property to that item differed depending on whether it was contrasted with less diverse or more diverse negative evidence.

Received September 18, 2018
Revision received February 11, 2019
Accepted February 17, 2019 ■