

What is typical about the typicality effect in category-based induction?

JONATHAN R. REIN, MICAH B. GOLDWATER, AND ARTHUR B. MARKMAN
University of Texas, Austin, Texas

Research on category-based induction has documented a consistent typicality effect: Typical exemplars promote stronger inferences about their broader category than atypical exemplars. This work has been largely confined to categories whose central tendencies are also the most typical members of the category. Does the typicality effect apply to the broad set of categories for which the ideal category member is considered most typical? In experiments with natural and artificial categories, typicality and induction-strength ratings were obtained for ideal and central-tendency exemplars. Induction strength was greatest for the central-tendency exemplars, regardless of whether the central tendency or the ideal was rated more typical. These results suggest that the so-called “typicality” effect is a special case of a more universal central-tendency effect in category-based induction.

To navigate the ever-changing world effectively, people must generalize from limited experience to a potentially infinite set of related but not-yet-experienced objects and events. One way to use this experience efficiently is to make inferences about entire categories, not just individual items. For instance, consider a homeowner who notices that the sparrows that live in her trees have been eating more of the new and improved birdseed that she purchased. She could reasonably infer not only that these specific sparrows will continue to prefer the new seed, but also that all sparrows—and possibly all birds—would, as well. This kind of predictive inference is a primary function of category knowledge (Markman & Ross, 2003; Osherson, Smith, Wilkie, López, & Shafir, 1990).

There are many factors that influence the strength of category-based property induction. These include the number of items known to have the property, the variability of the category one is generalizing to, the kind of information made salient by the property, and the causal relationships between categories and properties (Heit, 2000). In this article, we focus on how the item one is generalizing from affects its broader superordinate category. In particular, we examine how an item’s typicality—how good an example it is of its category—affects induction. In the example above, the homeowner may generalize from sparrows to all birds, because sparrows are highly representative of the bird category. In contrast, if she observes vultures enjoying the seed instead of sparrows, she will be much less likely to generalize to all birds because vultures are considered a fairly poor example of the bird category.

Rips (1975) was the first to document this typicality effect in category-based induction. In this study, participants were asked to imagine an island that contained a small set of bird or mammal species. They were told that a particular species had a contagious disease and were asked to esti-

mate the percentage of each other species that would share the disease. Estimates were significantly higher when the given instance was a typical category member, as measured by its distance from the superordinate category label in a multidimensional space. For example, when told that sparrows had the disease, people estimated that 32% of geese also had the disease. In contrast, they judged that only 17% of geese would have the disease when they were told that eagles had it. Sparrows and eagles are approximately equal in their similarity to geese, but sparrows had greater inductive strength because they are considered better examples of birds overall.

Another classic examination of category-based induction is Osherson et al.’s (1990) work on argument evaluation. These authors found that typicality has a substantial effect on judgments of inductive arguments. For example, participants were asked to choose which of the following is a stronger argument.

1. Robins have a higher potassium concentration in their blood than humans; therefore, all birds have a higher potassium concentration in their blood than humans.
2. Penguins have a higher potassium concentration in their blood than humans; therefore, all birds have a higher potassium concentration in their blood than humans.

Over 90% of people chose the robin argument.

This increase in induction strength for better category examples is a robust phenomenon. The typicality effect has been documented in multiple tasks, such as those just discussed, with multiple domains of category knowledge (Rothbart & Lewis, 1988) and for multiple kinds of subject populations, including children (López, Gelman, Gutheil, & Smith, 1992), indigenous Mayans in Guatemala (López,

Atran, Coley, Medin, & Smith, 1997), and Alzheimer's-disease patients (Smith, Rhee, Dennis, & Grossman, 2001). Although varied in several respects, these studies have one thing in common. They have all focused on a particular class of categories: what Markman and Stilwell (2001) refer to as "feature-based" categories.

Feature-based categories are those, such as *bird*, that are represented by their properties or dimension values. There have been several proposals of exactly how feature-based categories are represented, such as summary representations of average or characteristic values (Posner & Keele, 1968; Rosch & Mervis, 1975), the set of unique property combinations for each exemplar in the category (Medin & Shaffer, 1978; Nosofsky, 1986), and networks of causally related features (Rehder & Kim, 2006; Slovic, Love, & Ahn, 1998). For each of these, the category label denotes the set of features associated with that category; in other words, feature-based category representations contain information subordinate to, or part of, the category.

For these categories, an exemplar is considered to be a good or typical member of the category, to the extent that it has features characteristic of the category (Rosch & Mervis, 1975). These exemplars are closest to the central tendency of all category members. For example, swallows have many of the characteristic features of birds: They have wings, fly, build nests, have a common shape and size, and so on. They also do not have any features highly untypical of birds, as penguins do. These central-tendency exemplars are privileged in many ways, beyond being considered the best examples of their category and promoting the strongest induction. For example, these category members are classified faster (Rips, Shoben, & Smith, 1973) and with more certainty (Dale, Kehoe, & Spivey, 2007), brought to mind more readily (Mervis, Catlin, & Rosch, 1976), and learned more quickly (Rosch, Simpson, & Miller, 1976), than are less central exemplars.

Although feature-based categories have been the primary object of study in psychological research on concepts, there are other types of categories with different determinants of typicality. Role-governed categories, for instance, are qualitatively different from feature-based categories (Gentner & Kurtz, 2005; Markman & Stilwell, 2001). A role-governed category label refers to the role that an object plays in a larger relational system that connects multiple categories. Consider the category *predator*. In order for an object to qualify as a predator, it must play the first role in the relation *X hunts Y to satisfy a need*. Anything that plays this role is a predator, regardless of its feature composition. Lions and spiders are both predators, despite having drastically different features, because they both play this role. Although there are some features that many predators share (big teeth come to mind), these features are not relevant for an object's membership in the *predator* category.

Although they have been underrepresented in the research literature, role-governed categories are abundant. Of the 100 most common nouns, approximately half denote categories of roles or relations (Gentner & Kurtz, 2005). Classes of role-governed categories that have been

studied include verb-specific thematic roles (McRae, Ferretti, & Amyote, 1997), abstract coherent categories (Rehder & Ross, 2001), ad hoc categories (Barsalou, 1983), and goal-derived categories (Barsalou, 1985).

Goldwater, Stilwell, and Markman (2008, 2009) distinguished between role-governed categories and feature-based categories in a number of ways. Participants listed properties for sets of both kinds of categories. Barr and Caplan (1987) pointed out that there are two kinds of properties listed in such tasks. *Intrinsic* properties name characteristics that describe a category member; for example, "furry" is an intrinsic property of dogs. *Extrinsic* properties point outward to other objects; for example, "being owned by people" is an extrinsic property of dogs. Because role-governed categories are pieces of relational systems that bind concepts together, Goldwater et al. found that role-governed categories primarily elicit extrinsic properties—that is, properties that are about the connections among such bound concepts. In contrast, feature-based categories elicited primarily intrinsic properties. In addition, Goldwater et al. showed that these laboratory findings were relevant in predicting natural category use. The proportion of extrinsic properties listed predicted patterns of usage of the category labels to "tag" uploaded photographs by users of the popular photo-sharing Web site Flickr.com. These studies show that, although feature-based and role-governed categories lie in a continuum, both having features and relational content, the two are fundamentally different.

Considering the differences between feature-based and role-governed categories, one would expect differences in the nature of the best examples of these categories. Role-governed categories are not represented as bundles of correlated common features, but as elements that fit into a relation or serve a function. Consequently, a good example of a role-governed category is not an exemplar near the central tendency of the category, but rather an exemplar that best fits its relation or serves its function. Consider the category *diet food*. The central-tendency exemplar of a diet food is one with an average number of calories and a mediocre taste. In contrast, the best example of a diet food is one with zero calories and a great taste, even though this particular exemplar may not even exist. This kind of ideal exemplar is considered the most typical¹ for a variety of relational and goal-relevant categories (Barsalou, 1985; Borke, 1990; Davis & Love, 2008; Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003; Goldwater et al., 2008; Lynch, Coley, & Medin, 2000).

Extending these findings, participants in Goldwater et al. (2008, 2009) listed properties of both ideal and central-tendency exemplars of role-governed and feature-based categories. Ideal and central-tendency exemplars were constructed for every category from the five most frequent properties listed for the ideal and central-tendency conditions, excluding those characteristics listed for both. Then, another set of participants were asked to choose which exemplars would better illustrate the categories to someone who had no knowledge of them. Ideal exemplars were selected more for role-governed categories than for feature-based ones.

This dissociation in typicality between feature-based and role-governed categories raises an interesting question: In category-based induction that has been documented extensively with feature-based categories, is the typicality effect a result of the enhanced inductive strength of typical exemplars' typicality per se, or of their proximity to the central tendency? Studies of feature-based categories alone cannot resolve this question, because for these categories an exemplar's representativeness and its proximity to the central tendency are highly, if not perfectly, correlated. Fortunately, these are decoupled in role-governed categories for which the ideal, not the central tendency, is most typical.

There are two straightforward possibilities. If there is truly a typicality effect in category-based induction, central-tendency exemplars will promote stronger induction than will ideal exemplars for feature-based categories, whereas ideal exemplars will promote stronger induction than central-tendency exemplars for role-governed categories. In contrast, if the "typicality" effect is actually a central-tendency effect, central exemplars will promote stronger induction for both kinds of categories, whether or not they are more typical.

In this article, we evaluate the relative contribution of typicality and central tendency to category-based induction. In Experiment 1, participants make judgments about example goodness and induction strength for ideal and central-tendency exemplars of natural feature-based and role-governed categories. In Experiment 2, we manipulate the learning context of artificial categories to foster feature-based or role-governed representation, again measuring how idealness and central tendency influence example goodness and induction strength. These experiments will determine whether the typicality effect is a universal phenomenon for all kinds of categories, or simply a misnomer.

EXPERIMENT 1

Goldwater et al. (2008) documented several differences between common feature-based and role-governed categories. In one study, participants chose either an ideal or central-tendency exemplar as the best example to explain the category to someone unfamiliar with the category. As expected, people chose significantly more ideals to explain role-governed than feature-based categories. The present study is designed to replicate this dissociation in typicality judgments, and to evaluate whether typicality or central tendency is the main locus of induction strength.

Method

Participants. Forty-six University of Texas at Austin students participated for course credit or for payment of \$8 each.

Materials. All stimuli were adapted from Goldwater et al. (2008). There were 12 role-governed categories: *guest, job, game, predator, hobby, gift, drug, customer, home, author, friend, and pet*. Role-governed categories are pieces of relational systems. Verbs are represented by such relational structures (e.g., Dowty, 1979; Gentner, 1982). Not surprisingly, it seems that every role-governed category noun has a corresponding verb or verb phrase (see Goldwater et al., 2009, for an extended discussion of the relationship between verbs and role-governed categories). *Predator* relies crucially on verbs

like *hunt*, because the defining characteristic of a *predator* is that this individual is the first argument to the relation $X \text{ hunts } Y$. In addition to *predator*, a number of role-governed categories were selected for such a connection to a verb. The representations of *guest, author, gift, customer, home, and game* are dependent on *visit, write, give, shop, live, and play*. Here is a justification for the rest: *job and hobby* are role-governed, because classifying an activity as one or the other depends on the role it plays in your life (e.g., basketball is a job for some, but a hobby for many). A pet is an animal a person owns. Something is a drug on the basis of its ability to change one's mental state in some way. For example, industrial hemp, because it does not contain the psychoactive chemical THC found in other strains of cannabis, is not a drug. Clearly, people are classified as friends on the basis of the roles they play in your social life.

In addition to 12 role-governed categories, there were 12 feature-based categories: *television, chair, cell phone, fridge, truck, beer, Web site, shoes, knife, table, bicycle, and microwave*. These were all artifacts. Whereas natural kinds—particularly plants and animals—are more often used in studies of category-based induction (e.g., Osherson et al., 1990), artifacts were used because they are a stricter control for role-governed categories than natural kinds are. Because of their functions, artifact ideals are easily conceivable, whereas this is less true for natural kinds: For example, it is clearer what an ideal *television* is than what an ideal *turtle* is.

For each category, there was an ideal and central-tendency exemplar. Goldwater et al. (2009) obtained these exemplars by having separate groups of participants list the attributes of an ideal and average member of each category. Ideal exemplars are defined by the five most common attributes listed by the ideal group but not by the central-tendency group, and likewise for the central-tendency exemplars. There were 48 total exemplars. (See Appendix A for the entire set of categories and exemplars.)

Procedure. Participants were tested at individual computers. Each participant made example-goodness and induction-strength judgments for all 48 exemplars. Trials were blocked by judgment task. For the following task descriptions, we will use the category *predator* as an example.

The following instructions preceded the example-goodness block.

In this section of the study, we are interested in how people think of individual items as representative of the entire category to which the items belong. You will be presented with a series of items, one at a time. Each item has a list of five attributes. For each item, you will judge how good of an example of its category that item is. You will answer by selecting a point on a scale with your mouse. Although you'll see multiple members of each category, try to make each of your judgments independently from the rest.

On each trial of the exemplar-goodness task, this instruction was at the top of the screen: "Imagine a predator has the following properties." The five properties of the exemplar appeared below. For the ideal predator exemplar, these properties were "Smart," "Strong," "Agile," "Cunning," and "Camouflaged." Beneath the exemplar properties was the question "How good of an example of predators is this?" There was an 11-point scale at the bottom of the screen, with the lowest point labeled "poor example," the midpoint labeled "moderate example," and the highest point labeled "excellent example." Participants mouse-clicked a point on the scale.

The following instructions preceded the induction-strength block.

In this section of the study, we are interested in how people generalize from an individual item to the entire category to which the item belongs. You will be presented with a series of items, one at a time. Each item has a list of five attributes. For each item, you will be asked to imagine that the item has some additional attribute and to estimate the percentage of the rest of the category that also has that attribute. You will answer by

selecting a point on a scale with your mouse. Although you'll see multiple members of each category, try to make each of your judgments independently from the rest.

The induction-strength task was very similar to the example-goodness task. On each trial, participants were given an instruction along these lines: "Imagine that a predator with the following properties also has property X." "Property X" is an example of what is referred to in the induction literature as a *blank predicate*, in the sense that it has no meaningful content for the participant. This removes any role of background causal knowledge. The five properties of the exemplar were listed, followed by the question "What percentage of other predators have property X?" There was an 11-point scale at the bottom of the screen, with the lowest point labeled "0%," the midpoint labeled "50%," and the highest point labeled "100%."

Order of block presentation was counterbalanced. Exemplar order within a block was randomized, with two restrictions. First, the initial 24 trials in a block were exemplars of all 24 categories, as were the latter 24 trials. Second, within each of these sets, 12 of the exemplars were ideals, and 12 were central tendencies. Before each block, participants were instructed to make each judgment independent from the other judgments.

Results

The mean rating was calculated for each category type (feature-based vs. role-governed), exemplar type (central-tendency vs. ideal), and judgment task (goodness vs. induction), collapsing across participants and items for separate analyses. Figures 1 and 2 show participants' means for goodness and induction ratings. Table 1 contains the mean ratings for each category. Because the main question of interest is whether the relationship between category type and exemplar type is the same or different across tasks, we analyzed each task separately with 2 x 2 repeated measures ANOVAs on participant and item data. (For the interested reader, results from the full 2 x 2 x 2 ANOVAs are in Appendix B.) For each effect discussed below, the result of the participant analysis is first, followed by the item analysis. In order to generalize to the broader population of participants and items, treating both as random effects simultaneously, *Min F'* was also calculated for all effects (Clark, 1973).

Analyzing goodness ratings alone, there was no significant main effect of category type [$F(1,45) = 2.93$,

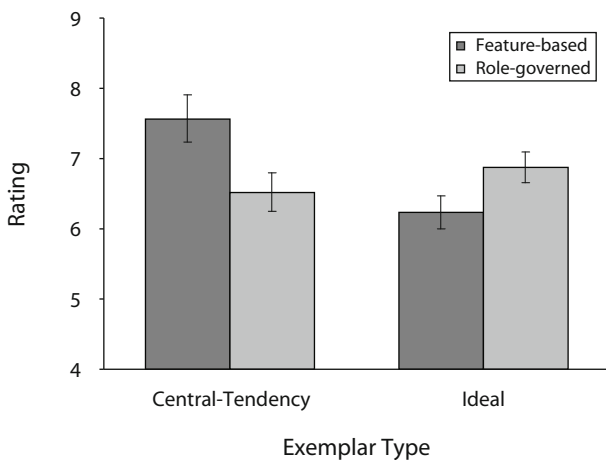


Figure 1. Experiment 1 example-goodness judgments. Error bars indicate ±1 SEM.

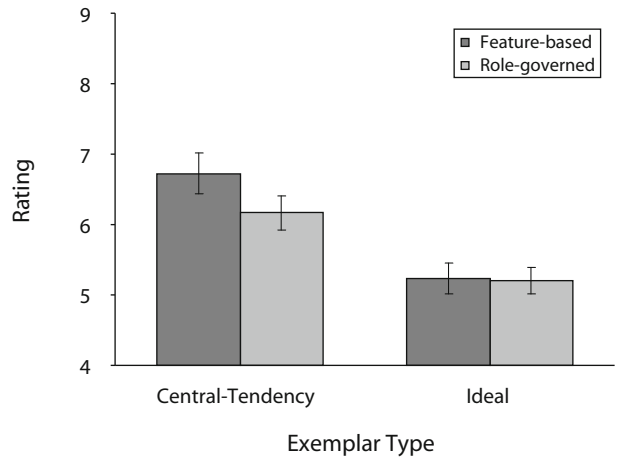


Figure 2. Experiment 1 induction-strength judgments. Error bars indicate ±1 SEM.

n.s.; $F(1,22) = 0.52$, *n.s.*; *Min F'*(1,30) = 0.44, *n.s.*]. There was no main effect of exemplar type by participants [$F(1,45) = 2.86$, *n.s.*]. This effect did reach significance by items [$F(1,22) = 5.71$, $p = .026$, $\eta^2 = .07$], though it was not significant across participants and items [*Min F'*(1,67) = 1.91, *n.s.*]. There was, however, a significant interaction between the category type and exemplar type [$F(1,45) = 18.47$, $p < .001$, $\eta^2 = .10$; $F(1,22) =$

Table 1
Example-Goodness and
Induction-Strength Ratings, by Category

	Goodness		Induction	
	Central	Ideal	Central	Ideal
Role Governed				
author	6.4	7.2	6.0	5.0
customer	4.3	6.5	5.0	4.7
drug	6.1	4.9	5.4	4.2
friend	8.2	8.3	6.0	5.7
game	7.6	6.7	6.8	5.5
gift	7.4	6.3	6.6	4.8
guest	6.3	7.1	6.1	5.6
hobby	5.7	7.1	5.7	5.9
home	7.4	6.4	7.0	4.4
job	5.1	6.8	6.1	4.4
pet	7.3	7.8	6.8	5.8
predator	6.9	7.7	6.4	6.3
<i>M</i>	6.6	6.9	6.1	5.2
Feature Based				
beer	8.0	4.4	7.3	4.3
bicycle	8.5	5.6	7.4	5.3
cell phone	7.5	5.9	6.8	5.1
chair	7.6	6.4	5.9	5.3
knife	7.6	6.9	6.4	5.8
microwave	7.6	6.7	6.7	5.4
refrigerator	7.8	7.0	7.0	5.8
shoes	6.7	6.0	6.0	4.6
table	7.1	6.5	5.8	5.7
television	7.6	6.2	7.2	4.9
truck	7.4	6.4	6.4	5.1
Web site	7.5	7.2	7.3	5.5
<i>M</i>	7.6	6.3	6.7	5.2

16.94, $p < .001$, $\eta^2 = .20$; $\text{Min } F'(1,57) = 8.84$, $p = .004$]. Central-tendency exemplars were considered better examples of feature-based categories ($M = 7.6$) than were role-governed categories ($M = 6.5$) [$F(1,45) = 19.64$, $p < .001$, $\eta^2 = .44$; $F(1,22) = 8.62$, $p = .008$, $\eta^2 = .28$; $\text{Min } F'(1,42) = 5.99$, $p = .02$]. In contrast, ideal exemplars were considered better examples of role-governed than of feature-based categories [$F(1,45) = 8.37$, $p = .006$, $\eta^2 = .19$], though this difference was only marginally significant by items [$F(1,22) = 3.59$, $p = .071$, $\eta^2 = .14$; $\text{Min } F'(1,41) = 2.51$, n.s.].

The induction-rating behavior was very different. There was a main effect of category type [$F(1,45) = 5.39$, $p = .025$, $\eta^2 = .02$], though this did not reach significance by items [$F(1,22) = 2.26$, n.s.; $\text{Min } F'(1,41) = 1.59$, n.s.]. There was also a main effect of exemplar type [$F(1,45) = 35.22$, $p < .001$, $\eta^2 = .27$; $F(1,22) = 68.27$, $p < .001$, $\eta^2 = .50$; $\text{Min } F'(1,67) = 23.23$, $p < .001$]. However, there was no significant interaction [$F(1,45) = 2.67$, n.s.; $F(1,22) = 3.12$, n.s.; $\text{Min } F'(1,62) = 1.44$, n.s.]. For feature-based categories, central-tendency exemplars were considered stronger bases for induction than were ideals [$F(1,45) = 21.83$, $p < .001$, $\eta^2 = .48$; $F(1,22) = 46.40$, $p < .001$, $\eta^2 = .68$; $\text{Min } F'(1,67) = 14.84$, $p < .001$]. For role-governed categories, central-tendency exemplars ($M = 6.2$) also had higher induction ratings than ideals did ($M = 5.2$) [$F(1,45) = 26.18$, $p < .001$, $\eta^2 = .58$; $F(1,22) = 12.38$, $p = .002$, $\eta^2 = .36$; $\text{Min } F'(1,43) = 8.41$, $p = .006$].

Overall, there was a positive correlation between goodness ratings and induction ratings for individual items [$r(48) = .68$]. It is informative to break this correlation down by category type. There was a significant correlation for role-governed categories [$r(24) = .47$]. There was also a significant correlation for feature-based categories [$r(24) = .89$, $p < .001$; $\text{Min } F'(1,103) = 6.52$, $p = .01$]. As expected, the correlation between goodness and induction ratings was stronger for feature-based than for role-governed categories ($z = 2.95$, $p = .003$).

Discussion

Participants' judgments of goodness of example for the different exemplar types mirrored previous work. Ideal exemplars were considered to be much better examples of role-governed categories than feature-based categories were. In contrast, central-tendency exemplars were considered better examples of feature-based than of role-governed categories were.

A different pattern was observed for induction strength. Participants rated central-tendency exemplars as promoting stronger induction for both feature-based and role-governed categories. For feature-based categories, this replicates the canonical typicality effect. For role-governed categories, however, the most representative exemplars do not promote the strongest induction. This suggests that representativeness per se is not in fact responsible for the typicality effect.

Although not central to the main point of this article, it is interesting to note that induction ratings were lower for role-governed categories than for feature-based categories overall. This may be because categories represented primar-

ily with relational properties exhibit less within-category similarity (Gentner & Kurtz, 2005). Role-governed category members all share an identical role, but typically they do not all share a bundle of correlated attributes, as feature-based category members do. This within-category coherence has been shown to influence inductive strength (Patalano, Chin-Parker, & Ross, 2006). Care should be taken in drawing conclusions from this finding, however, because it was not statistically reliable in the items analysis.

Experiment 1 provides preliminary evidence that category-based induction is actually influenced principally by exemplars' proximity to their categories' central tendency. However, this study used natural categories and therefore leaves the causality underlying this effect ambiguous. Role-governed categories differ in several ways from feature-based categories (Goldwater et al., 2008), and any one of these might have overridden the true typicality effect. Experiment 2 addresses this ambiguity by using an artificial category learning task.

EXPERIMENT 2

In this study, we adopt a method used by Barsalou (1985) to foster feature-based or role-governed representation of identical category information simply by manipulating the learning context. Participants learned to classify two categories of teachers, with each teacher exemplar defined by five spare-time activities in which they engage. Each category had a defining activity. For one category, all of the exemplars had the activity *reads the newspaper*, whereas all of the members of the other category had the activity *jogs*. Individual exemplars varied in how often they engaged in this defining activity: *daily*, *weekly*, or *monthly*. In addition to this defining activity, each of the categories had three characteristic activities that predicted membership in that category but not in the other. For instance, one category's members would tend to have the activities *invests in stocks*, *writes poetry*, and *goes to the movies*, whereas members of the other category would generally have *plays chess*, *renovates houses*, and *cooks Mediterranean food*. Individual exemplars varied in how many characteristic activities they had: three, two, or one. Critically, one group of participants learned these categories with role-relevant labels: "current events teachers" and "physical education teachers" who read the newspaper and jog, respectively. Another group of participants learned to distinguish the same exemplars under role-irrelevant labels: "Q programming language teachers" and "P programming language teachers."

Unlike the natural categories in Experiment 1, in which idealness and central tendency were mutually exclusive, this design manipulates both independently. Barsalou (1985) found that each has an influence on judgments of exemplar goodness, depending on label relevance. When participants learned the categories with role-relevant labels, the frequency of an exemplar's performing the defining activity determined how good an example it was. Physical education teachers who jogged daily were considered better examples than their peers who jogged monthly. In other words, ideal exemplars that best fit their role had

the highest typicality. For these role-relevant learners, an individual exemplar's number of characteristic features had little to no effect.

The relative influence of ideals and central tendency was reversed for role-irrelevant learners. An exemplar's frequency of the defining activity had no effect on category-goodness judgments. Jogging has no obvious impact on how well Q programming language teachers fill their functional role, so jogging daily does not make one any better a programming teacher than jogging monthly does. In contrast, the number of an exemplar's characteristic activities did influence goodness judgments. Without functional information, participants relied on statistical similarity. The programming teachers were essentially feature-based categories. As has been documented extensively, the most typical exemplars for these categories are those closest to the central tendency.

Experiment 2 replicates Barsalou's (1985) work on typicality, extending it to category-based induction. As in Experiment 1, if an exemplar's representativeness of its category is truly the determinant of induction, induction-strength ratings should map onto example-goodness ratings. If, however, exemplars' proximity to the category-central tendency is the determinant of induction, exemplars with more characteristic features should promote stronger induction, even if they are not considered most representative of their category.

Method

Participants. Sixty-five University of Texas students participated for course credit or payment of \$8 each; 32 were in the role-irrelevant condition, 33 in the role-relevant condition.

Materials. Each category consisted of nine exemplars. For each level of the defining activity (*daily*, *weekly*, and *monthly*) there was an exemplar with each level of characteristic activities (three, two, and one). The three characteristic activities were equally diagnostic of category membership, with each appearing in six of the nine exemplars. Every exemplar had 1–3 filler activities, yielding 5 activities per exemplar. There were 18 such filler activities, with each activity appearing in a single member of each category. For example, one category might have the defining activity *reads the newspaper* and the characteristic activities *plays chess*, *renovates houses*, and *cooks Mediterranean food*. In this category, a high-defining/high-central-tendency exemplar would have the following properties: *reads the newspaper daily*, *plays chess*, *renovates houses*, *cooks Mediterranean food*, and *goes fishing*. This last property is a filler activity that is true of only this exemplar and one other exemplar in the contrast category. A low-defining/low-central-tendency exemplar would have the following properties: *reads the newspaper daily*, *renovates houses*, *plays the drums*, *rides a motorcycle*, and *goes to yard sales*. This exemplar has 3 filler activities.

Procedure. Participants were tested at individual computers. They were told to imagine that they worked for a personnel agency, where they were learning to use information about individuals' spare-time activities to predict the course that they would be best suited to teach. In the role-irrelevant condition, they were told that they would be distinguishing between Q programming language teachers and P programming language teachers. In the role-relevant condition, they were told that they would be distinguishing between current events teachers and physical education teachers.

Prior to the classification phase, participants were familiarized with the exemplars from each category. During this familiarization phase, a category label appeared at the top of the screen, with the five activities describing all nine exemplars presented below. Partici-

pants studied the exemplars for a minimum of 2 min, then viewed all of the exemplars from the other category. Order of category presentation was randomized.

On each trial of the classification phase, an instruction appeared at the top of the screen, telling participants to "Choose the better teacher of [name of course]," with each course label appearing in half of the trials. An exemplar from each category appeared below this instruction, one on the left and one on the right. Categories' position was randomized across trials. Participants pressed the "Z" or "/" button on the keyboard to select the left or right exemplar. After each classification choice, "Correct" or "Incorrect" appeared at the bottom of the screen, the correct exemplar was highlighted by a white box, and a pleasant tone or harsh buzzing sound played. Participants studied the correct exemplar for 8 sec, then were presented with a single activity and were asked whether it had been true of the correct exemplar. On half of the trials, this activity was true of the correct exemplar; on the other half it was true of the incorrect exemplar. Because participants could correctly classify all of the exemplars using only the defining activity, the memory probe and the familiarization phase were included to promote complete knowledge of the categories, including the characteristic activities. There were 54 total classification trials, 3 each for all 18 exemplars.

Following classification, participants provided example-goodness and induction-strength judgments for all 18 exemplars. As in Experiment 1, these judgments were blocked, with block order counterbalanced across participants. Exemplar presentation within blocks was randomized. The presentation of each judgment trial was identical to that in Experiment 1, except that induction judgments referred to "activity X" instead of "property X."

Results

The mean of participants' ratings across exemplars for each level of idealness on the defining activity (*daily*, *weekly*, and *monthly*), characteristic activities (three, two, and one), judgment task (goodness vs. induction), and label relevance (irrelevant vs. relevant) was calculated. Figures 3 and 4 show the means across participants for goodness and induction ratings, respectively. As in Experiment 1, we analyzed the data for each task in each labeling condition separately with 3×3 repeated measures ANOVAs. (The full 4-way ANOVA table can be found in Appendix B.)

In the case of example-goodness judgments following role-irrelevant learning, more characteristic activities—that is, greater central tendency—led to higher ratings [$F(2,62) = 32.74, p < .001, \eta^2 = .25$]; this was a significant linear trend [$F(1,31) = 40.81, p < .001, \eta^2 = .25$]. There was no effect of ideal level overall [$F(2,62) = 1.76, n.s.$]. There was, however, an interaction between central-tendency level and ideal level [$F(4,124) = 4.57, p = .002, \eta^2 = .04$]. This effect appears to be due to the exceptionally low ratings for exemplars, with only one characteristic activity and infrequent performance of the defining activity.

Example-goodness judgments following role-relevant learning were much different. As in the role-irrelevant condition, ratings increased as central tendency increased [$F(2,64) = 11.84, p < .001, \eta^2 = .06$]; this was a significant linear trend [$F(1,31) = 15.8, p < .001$]. However, this effect was substantially smaller than in the role-irrelevant condition, as evidenced by the relative effect sizes, as well as by a significant two-way interaction between central tendency and label relevance in the context of a three-

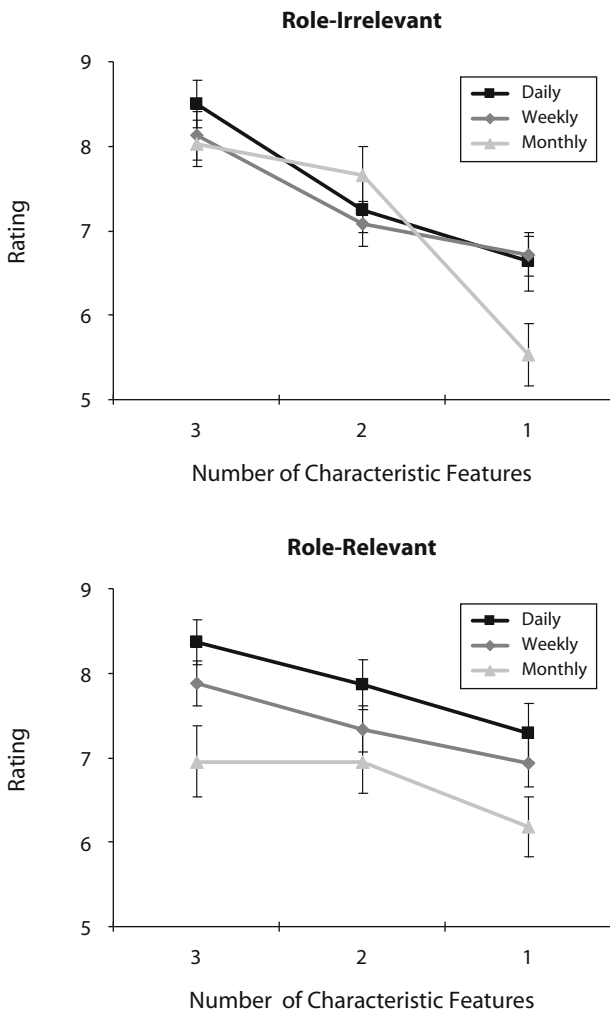


Figure 3. Experiment 2 example-goodness judgments. Error bars indicate ± 1 SEM.

way (central tendency \times ideal level \times label relevance) ANOVA on goodness judgments alone [$F(2,126) = 5.02, p = .008, \eta^2 = .04$]. There was also an effect of ideal level, with more ideal exemplars rated as better examples [$F(2,64) = 6.09, p = .004, \eta^2 = .09$]; this was a significant linear trend [$F(1,31) = 6.65, p = .015, \eta^2 = .09$]. There was no significant interaction between these variables [$F(4,128) = 0.75, n.s.$]. This increased influence of ideals on example-goodness judgment following role-relevant learning is a replication of Barsalou's (1985) findings.

Induction-strength judgments following role-irrelevant learning were similar to example-goodness judgments. Greater central tendency led to higher induction-strength ratings [$F(2,62) = 6.40, p = .003, \eta^2 = .07$]; this was a significant linear trend [$F(1,32) = 7.16, p = .012, \eta^2 = .06$]. There was no effect of ideal level [$F(2,62) = 1.15, n.s.$], nor was there any interaction between the two [$F(4,124) = 0.91, n.s.$].

In contrast, induction-strength judgments following role-relevant learning did not correspond to example-goodness ratings. As in the role-irrelevant condition,

increased central tendency led to increased induction strength [$F(2,64) = 9.94, p < .001, \eta^2 = .11$]; this was a significant linear trend [$F(1,32) = 13.15, p = .001, \eta^2 = .10$]. There was no effect of ideal level [$F(2,64) = 1.01, n.s.$]. Although the low-central-tendency, high-ideal data point appears somewhat higher, there was no statistically significant interaction between central tendency and ideal level [$F(4,128) = 1.22, n.s.$].

Discussion

In the role-irrelevant condition, there was no obvious relationship between the categories' functions and their constituent activities. Consequently, participants represented the teacher groups as feature-based categories described by the distribution of activities. As with other feature-based categories, central-tendency exemplars were considered the best examples. The level of the defining activity had little effect on example goodness, because it had no bearing on the statistics of the category, nor could it be construed as a role-relevant ideal dimension.

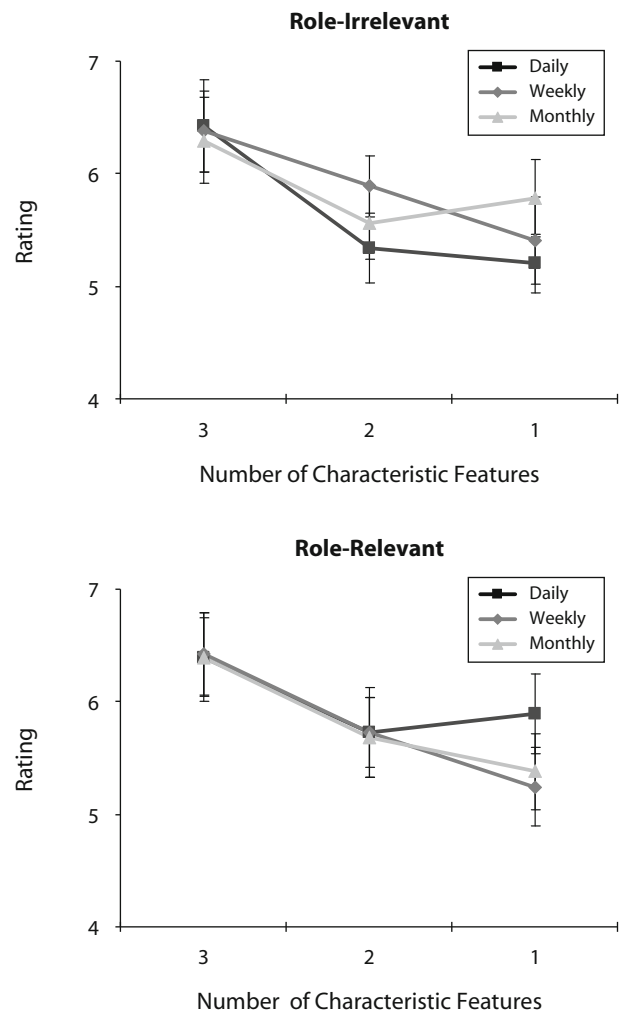


Figure 4. Experiment 2 induction-strength judgments. Error bars indicate ± 1 SEM.

Judgments of induction strength mirrored example goodness in the role-irrelevant condition. Proximity to central tendency increased induction strength, whereas idealness had no effect on these judgments. This is another replication of the traditional typicality effect observed in feature-based categories.

In the role-relevant condition, the categories' functions and the relationship between those functions and the defining activities were salient. As a result, the teacher groups were represented as role-governed categories. As with other role-governed categories, ideal exemplars were considered the best examples. The level of central tendency also had an effect on example goodness. This is not surprising, considering the design's emphasis—through the familiarization phase and memory probes—on learning all category information. Nevertheless, this effect was smaller than was the effect of ideals in this condition, and substantially smaller than the effect of central tendency in the role-irrelevant condition.

In contrast to example-goodness judgments, participants considered only central tendency during judgments of induction strength in the role-relevant condition. There was no effect of idealness on induction, for either ideals or anti-ideals. The influence of central-tendency information on induction for role-relevant categories was remarkably similar to that for role-irrelevant categories, despite the divergence in example-goodness judgments.

GENERAL DISCUSSION

In experiments with natural and artificial categories, we replicated previous work showing that role-governed categories were represented better by ideal exemplars than were feature-based categories, which were better represented by central-tendency exemplars (Goldwater et al., 2008). Consistent with the canonical "typicality" effect, central-tendency exemplars also promoted the strongest induction for feature-based categories. Perhaps surprisingly, there was no such typicality effect in induction for role-governed categories. Instead, exemplars closest to the category central tendency provided the greatest induction strength.

It is possible that there is a true typicality effect in category-based induction for feature-based categories and a central-tendency effect for role-governed categories. However, a more parsimonious explanation of these findings is that central-tendency exemplars are inductively privileged for all category types. The typicality effect for feature-based categories appears to be a special case of this more general phenomenon. This universality in category-based induction also places boundary conditions on the feature-based/role-governed dissociation. Although these two kinds of categories differ in many ways (Goldwater et al., 2008; Markman & Stilwell, 2001), there is at least one cognitive process in which they behave identically.

Although this finding of a central-tendency effect instead of a typicality effect is novel, it is not actually incompatible with existing models of category-based induction. This is precisely because many of these models were designed to accommodate induction behavior for natural-

kind categories, whose central tendencies are most typical. For example, under Osherson et al.'s (1990) model, the strength of an induction from the category *robins* to the category *birds* is a function of *robins*' "coverage" of *birds*. *Robin* is on average very close to other categories in the multidimensional space of *birds* (i.e., more similar to other birds). In other words, *robins* covers the space of *birds* more broadly than does *penguins*. Sloman's (1993) model embodies this idea of coverage through exemplars' and categories' feature overlap in a connectionist architecture. Again, central tendencies have greater induction strength as premises, because they have the characteristic features of the category, and therefore have substantial overlap with the category's features. Bayesian models of induction (e.g., Heit, 1998; Kemp & Tenenbaum, 2009) can also account for central-tendency advantages, although they could also account for ideal advantages, depending on the nature of the hypothesis space.

All of these models are consistent with the experiments reported here, because they all perform some kind of rough statistical calculation in which the central tendency has a special status. The central tendency lies at the statistical center of the category distribution, and is, therefore, consistent with the most hypotheses about the property's distribution within the category. Critically, this statistical relationship holds for feature-based *and* role-governed categories.

Another open question is how this simple central-tendency-based induction process relates to experts' more sophisticated induction. Coley, Medin, and Atran (1997) have found that many experts' inductive reasoning has a character distinct from novices'. For instance, tree experts, such as landscapers and park maintenance workers, consider an item to be a good example of the *tree* category if it has more extreme values on ideal dimensions such as height and weediness (Lynch et al., 2000). However, when told that two trees have two different diseases, the relative idealness of the trees has no influence on which disease the experts expect to be present in all trees. Interestingly, the relative central tendency also has no influence. Instead, tree experts rely on causal ecological knowledge, like a tree's relative susceptibility to disease, and which trees are likely to be nearby (Proffitt, Coley, & Medin, 2000). Similarly, indigenous Maya, who are experts in their native ecology, consider turkeys to be the best example of birds because of ideal features like large size, obvious markings, and being a plentiful food source (Atran, 1999). Nevertheless, their induction about the presence of disease in birds is based on ecological knowledge, not on ideals or central tendency (Bailenson, Shum, Atran, Medin, & Coley, 2002).

There are three principal differences between this work on experts' ideal-represented categories and the work reported in this article. First, the experts have a great deal of functional knowledge, but trees and birds are still natural-kind categories. It is therefore not obvious where they fall in regard to the feature-based/role-governed distinction. Second, as Coley et al. (1997) have noted, experts have a rich set of causal knowledge that novices do not have, even for familiar everyday categories. Third, and perhaps most

critically, when induction is performed with nonblank predicates, such as diseases, this vast knowledge can be used for complex causal reasoning; this is not possible for blank predicates unrelated to this knowledge base. It is possible that experts exhibit the central-tendency effect for blank predicates, just as novices have been shown to reason causally about natural categories and nonblank predicates (Smith, Shafir, & Osherson, 1993). Because the nature of our knowledge about role-governed categories is different from other kinds of categories (Goldwater et al., 2008, 2009), role-governed categories may show a unique pattern in supporting inductive inferences that make use of such background knowledge.

In addition to induction per se, the present results also inform our understanding of the relationship between categorization and category-based induction. Considering the overlap in category knowledge used in both behaviors, it is tempting to think of them as essentially the same process. In some tasks, categorization and induction behavior are remarkably similar and both can be predicted by a single similarity-based model (Heit & Hayes, 2008; Sloutsky & Fisher, 2004). In other tasks, however, these two behaviors are dissociated. Murphy and Ross (2005) showed that certainty of classification in the presence of contrast categories affects categorization, but not induction. In their work with American college students and indigenous Maya, Coley, Medin, and Atran (1997) showed that the basic level of categorization differed according to cultural expertise but that the genus level was inductively privileged for both groups. The experiments reported here show yet another way in which categorization and induction are dissociable.

CONCLUSION

The typicality effect in category-based induction has generally been documented with feature-based categories. Examining induction for role-governed categories reveals that this effect is not really about typicality at all. Instead, central-tendency exemplars promote the strongest induction, regardless of their typicality. This dissociation between typicality and induction sets limits on the feature-based/role-governed distinction and provides clues to the universal processes underlying category-based induction.

AUTHOR NOTE

This research was supported by NIMH Grant R01-MH0778 and by a fellowship to the IC² Institute given to A.B.M. We thank Lisa Grimm, Jeff Laux, Grant Loomis, Ross Otto, and Hunt Stilwell for their assistance with this project. We also thank Dustin Calvillo, Evan Heit, and two anonymous reviewers for their comments and suggestions on an earlier draft of this article. Correspondence concerning this article should be addressed to A. B. Markman, Department of Psychology, University of Texas, 1 University Station A8000, Austin, TX 78712 (e-mail: markman@psy.utexas.edu).

REFERENCES

- ATRAN, S. (1999). Itzaj Maya folkbiological taxonomy. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 119-204). Cambridge, MA: MIT Press.
- BAILENSEN, J. N., SHUM, M. S., ATRAN, S., MEDIN, D. L., & COLEY, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, **84**, 1-53.
- BARR, R. A., & CAPLAN, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition*, **15**, 397-418.
- BARSALOU, L. W. (1983). Ad hoc categories. *Memory & Cognition*, **11**, 211-227.
- BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-654.
- BORKENAU, P. (1990). Traits as ideal-based and goal-derived social categories. *Journal of Personality & Social Psychology*, **58**, 381-396.
- CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.
- COLEY, J. D., MEDIN, D. L., & ATRAN, S. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. *Cognition*, **64**, 73-112.
- DALE, R., KEHOE, C., & SPIVEY, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, **35**, 15-28.
- DAVIS, T., & LOVE, B. C. (2008). How goals shape category acquisition: The role of contrasting categories. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 101-106). Austin, TX: Cognitive Science Society.
- DOWTY, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht: Kluwer Publications.
- GENTNER, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought, & culture* (Vol. 2, pp. 301-334). Hillsdale, NJ: Erlbaum.
- GENTNER, D., & KURTZ, K. J. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151-175). Washington, DC: American Psychological Association.
- GOLDSTONE, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, **24**, 608-628.
- GOLDSTONE, R. L., STEYVERS, M., & ROGOSKY, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, **31**, 169-180.
- GOLDWATER, M. B., STILWELL, C. H., & MARKMAN, A. B. (2008). *The ideal representation of role-governed categories*. Poster presented at the 30th Annual Meeting of the Cognitive Science Society, Washington, DC.
- GOLDWATER, M. B., STILWELL, C. H., & MARKMAN, A. B. (2009). *The empirical case for role-governed categories*. Manuscript in preparation.
- HEIT, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248-274). Oxford: Oxford University Press.
- HEIT, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, **7**, 569-592.
- HEIT, E., & HAYES, B. K. (2008). Predicting reasoning from visual memory. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 83-88). Austin, TX: Cognitive Science Society.
- KEMP, C., & TENENBAUM, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, **116**, 20-58.
- LÓPEZ, A., ATRAN, S., COLEY, J. D., MEDIN, D. L., & SMITH, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, **32**, 251-295.
- LÓPEZ, A., GELMAN, S. A., GUTHEIL, G., & SMITH, E. E. (1992). The development of category-based induction. *Child Development*, **63**, 1070-1090.
- LYNCH, E. B., COLEY, J. D., & MEDIN, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, **28**, 41-50.
- MARKMAN, A. B., & ROSS, B. H. (2003). Category use and category learning. *Psychological Bulletin*, **129**, 592-613.
- MARKMAN, A. B., & STILWELL, C. H. (2001). Role-governed category

- ries. *Journal of Experimental & Theoretical Artificial Intelligence*, **13**, 329-358.
- MCRAE, K., FERRETTI, T. R., & AMYOTE, L. (1997). Thematic roles as verb-specific concepts. *Language & Cognitive Processes*, **12**, 137-176.
- MEDIN, D. L., & SHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
- MERVIS, C. B., CATLIN, J., & ROSCH, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, **7**, 283-284.
- MURPHY, G. L., & ROSS, B. H. (2005). The two faces of typicality in category-based induction. *Cognition*, **95**, 175-200.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- PATALANO, A. L., CHIN-PARKER, S., & ROSS, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory & Language*, **54**, 407-424.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.
- PROFFITT, J. B., COLEY, J. D., & MEDIN, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 811-828.
- REHDER, B., & KIM, S. W. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 659-683.
- REHDER, B., & ROSS, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1261-1275.
- RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.
- RIPS, L. J., SHOBBEN, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, **12**, 1-20.
- ROSCH, E., & MERVIS, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.
- ROSCH, E., SIMPSON, C., & MILLER, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 491-502.
- ROTHBART, M., & LEWIS, S. (1988). Inferring category attributes from exemplar attributes: Geometric shapes and social categories. *Journal of Personality & Social Psychology*, **55**, 861-872.
- SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, **25**, 231-280.
- SLOMAN, S. A., LOVE, B. C., & AHN, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.
- SLOUTSKY, V. M., & FISHER, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, **133**, 166-188.
- SMITH, E. E., RHEE, J., DENNIS, K., & GROSSMAN, M. (2001). Inductive reasoning in Alzheimer's disease. *Brain & Cognition*, **47**, 494-503.
- SMITH, E. E., SHAFIR, E., & OSHERSON, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, **49**, 67-96.

NOTE

1. "Typicality" has generally been treated in the literature as synonymous with "representativeness." Since Rosch's early work, both of these constructs have traditionally been assessed by goodness-of-example ratings. As a result, researchers describe ideal examples as most typical when they are rated as the best example of their category, even though they are not typical in the "average" sense. Following these conventions, we use "typicality," "representativeness," and "example goodness" interchangeably, while emphasizing that the third is best thought of as an operationalization of the first two.

APPENDIX A

Ideal and Central Tendency Characteristic Sets Used in Experiment 1

Role-Governed Categories

Author

Ideal: good writer, funny, books aren't too long, clear, relatable

Central tendency: middle-aged, imaginative, intellectual, reclusive, large vocabulary

Customer

Ideal: has a lot of money, polite, friendly, nice, knows what they want

Central tendency: female, impatient, has money, demanding, buying something

Drug

Ideal: no side effects, inexpensive, not addictive, legal, not impairing

Central tendency: cures, alters mind, painkiller, addictive, helpful

Friend

Ideal: understanding, intelligent, sympathetic, able to keep secrets, compatible

Central tendency: nice, helpful, kind, always there for you, fun

Game

Ideal: interesting, challenging, active, changes frequently, easy to learn

Central tendency: competitive, rules, athletic, winners, losers

Gift

Ideal: practical, something recipient wants, liked by recipient, handmade, reusable

Central tendency: wrapped, card, small, on special occasions, on birthdays

Guest

Ideal: clean, courteous, has manners, fun, unobtrusive

Central tendency: family, friend, kind, dressed up, invited

Home

Ideal: looks good, good location, big, comfortable, pool

Central tendency: has a kitchen, where family is, roof, windows, made of brick

APPENDIX A (Continued)

Hobby

Ideal: inexpensive, social, intellectually stimulating, exciting, healthy

Central tendency: time consuming, relaxing, done alone, requires a skill, done regularly

Job

Ideal: high paying, flexible hours, good boss, good coworkers, fun

Central tendency: boring, long hours, low wages, make money, time consuming

Pet

Ideal: playful, easy to care for, loyal, obedient, friendly

Central tendency: furry, lovable, loving, soft, fun

Predator

Ideal: smart, strong, agile, cunning, camouflaged

Central tendency: mean, sharp teeth, big, has claws, hungry

Feature-Based Categories
Beer

Ideal: tastes good, inexpensive, nonfattening, healthy, flavorful

Central tendency: bottled or canned, carbonated, yellow, alcoholic

Bicycle

Ideal: comfortable, fast, looks good, cheap, durable

Central tendency: handlebars, seat, made of metal, pedals, has reflectors

Cell Phone

Ideal: camera, durable, easy to use, inexpensive, has lots of memory

Central tendency: screen, key pad, ringtones, buttons, flip phone

Chair

Ideal: adjustable, reclining, wheels, leather, rolls around/has wheels

Central tendency: wooden, four legs, backrest, arms, seat

Fridge

Ideal: spacious, energy efficient, goes well with room, looks good, has a water dispenser

Central tendency: has a freezer, stores food, white, has two doors, has an automatic light inside

Knife

Ideal: easy to hold, comfortable, cuts well, durable, lightweight

Central tendency: dangerous, metal blade, silver blade, shiny, black handle

Microwave

Ideal: quiet, fast, energy efficient, inexpensive, easy to use

Central tendency: black, heats food, buttons, turning plate, light inside

Shoes

Ideal: look good, inexpensive, match everything, cute, last a long time

Central tendency: laces, rubber, white, soles, leather

Table

Ideal: sturdy, looks good, strong, adjustable, big

Central tendency: four legs, flat top, round or rectangular, glass top

Television

Ideal: clear reception, high definition, big, lightweight, inexpensive

Central tendency: black, color picture, square, remote control, screen

Truck

Ideal: fuel efficient, powerful, carries a lot, roomy, strong

Central tendency: four wheels, bed, gas guzzling, two door, loud

Web Site

Ideal: easy to use, looks good, useful, entertaining, fast loading

Central tendency: links, colorful, pictures, ads, search

APPENDIX B

Table B1
Experiment 1 Subjects Analysis ANOVA Table

Effect	<i>F</i>	<i>df</i>	<i>p</i>
Judgment task	22.75	1,45	<.001
Category type	10.64	1,45	.002
Exemplar type	19.482	1,45	<.001
Judgment task × category type	0.25	1,45	.619
Judgment task × exemplar type	5.42	1,45	.025
Category type × exemplar type	21.53	1,45	<.001
Judgment task × category type × exemplar type	4.7	1,45	.035

Table B2
Experiment 1 Items Analysis ANOVA Table

Effect	<i>F</i>	<i>df</i>	<i>p</i>
Judgment task	138.01	1,22	<.001
Category type	2.262	1,22	.147
Exemplar type	29.12	1,22	<.001
Judgment task × category type	0.309	1,22	.584
Judgment task × exemplar type	20.47	1,22	<.001
Category type × exemplar type	12.057	1,22	.002
Judgment task × category type × exemplar type	12.591	1,22	.002

Table B3
Experiment 2 ANOVA Table

Effect	<i>F</i>	<i>df</i>	<i>p</i>
Judgment task	46.522	1,63	<.001
Idealness	5.132	2,126	.007
Central tendency	41.506	2,126	<.001
Label relevance	0.038	1,63	.846
Judgment task × idealness	6.061	2,126	.003
Judgment task × central tendency	5.675	2,126	.004
Judgment task × label relevance	0.004	1,63	.949
Idealness × central tendency	1.445	4,252	.22
Idealness × label relevance	3.491	2,126	.033
Central tendency × label relevance	1.802	2,126	.169
Judgment task × idealness × central tendency	3.192	4,252	.014
Judgment task × idealness × label relevance	0.655	2,126	.521
Judgment task × central tendency × label relevance	2.872	2,126	.06
Idealness × central tendency × label relevance	0.702	4,252	.591
Judgment task × idealness × central tendency × label relevance	2.572	4,252	.038

(Manuscript received June 1, 2009;
revision accepted for publication October 9, 2009.)