MEMORY, 1996, *4* (4), 413–451

The Instantiation Principle in Natural Categories

Evan Heit
University of Warwick, UK

Lawrence W. Barsalou
University of Chicago, USA

According to the instantiation principle, the representation of a category includes detailed information about its diverse range of instances. Many accounts of categorisation, including classical and standard prototype theories, do not follow the instantiation principle, because they assume that detailed, exemplar-level information is filtered out of category representations. Nevertheless, the instantiation principle can be implemented in a wide class of models, including both exemplar and abstraction models. To assess the instantiation principle empirically, a parameter-free exemplar-based model of instantiation was applied to typicality judgments for 16 simple categories (e.g. *mammal*, *beverage*) and 14 complex categories (e.g. *dangerous mammal*) in four superordinates (*animal*, *food*, *small animal*, *dangerous animal*). Across three studies, the model did an excellent job of predicting mean typicality judgments (correlations generally above 0.9) and a good job of predicting standard deviations (fits generally from 0.6 to 0.9). In Study 3, predicting the skew of typicality distributions was successful as well (a fit of 0.87), and dropping atypical exemplars from the simulations degraded prediction. All of these results support the instantiation principle, indicating that subjects incorporate detailed information about category instances into their representations of categories.

Requests for reprints should be sent to Evan Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, UK. Email: E.Heit@warwick.ac.uk

A report on Study 3 was presented to the Experimental Psychology Society conference, at University College London, in January 1996. A preliminary report of Studies 1 and 2 was presented to the Cognitive Science Society conference, at Indiana University, in July 1992. This research was supported by a National Institute of Mental Health grant (1 F32 MH10069) to Evan Heit, by an Army Research Institute contract to Lawrence Barsalou (MDA 903-90-K-0112), by a National Science Foundation grant to Lawrence Barsalou (SBR-9421326), and by a National Science Foundation grant to Douglas Medin (91-10245). The authors are grateful to Dedre Gentner, James Hampton, Douglas Medin, Lance Rips, Brian Ross, Edward E. Smith, Jack Stecher, and Edward Wisniewski for comments on this research.

INTRODUCTION

What is the relation between what a person knows about a natural category and what is known about its instances? For example, what is the relation between knowledge of *mammal* and its instances, such as *dog*, *monkey*, and *human*? One possibility is that knowledge about a category represents a summary or reduction about what is known about the instances. By this account, the representation of *mammal* might contain information about what various mammals have in common, or at least what is highly frequent across them. This account highlights the ‘‘cognitive economy’’ function of categories proposed by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976). Categories are useful because they allow us to ignore idiosyncratic information about particular instances, establishing more economical representations of general features instead.

Two particularly well-known kinds of categorisation models implement the cognitive economy principle: classical models and prototype models (for reviews, see Barsalou & Hale, 1993, Hampton, 1993, and Smith & Medin, 1981). In classical models, a category is represented by a set of necessary and sufficient conditions. Typically, these models discard idiosyncratic information about instances in the process of establishing definitions. Likewise, standard prototype models typically discard idiosyncratic information and retain only high-frequency attributes in prototypes. There are numerous other ways to create category representations that also provide some reduction of information. For example, some connectionist models of categorisation essentially learn to form prototype-like summaries of general features (Nosofsky, 1992).

The instantiation principle that we explore here stands in contrast to the cognitive economy principle. According to this principle, knowledge about a general category reflects a great deal of detailed information about the diverse range of its instances. For example, the representation of *mammal* contains detailed information about particular instances such as *dog* and *monkey*.

Exemplar models of categorisation strongly embody the instantiation principle, because they assume that detailed information about instances enters into categorisation processes (Estes, 1994; Heit, 1992, 1994; Lamberts, 1994, 1995; Medin & Schaffer, 1978; Nosofsky, 1984). These models share the common assumption that, to evaluate whether stimulus x belongs to category A, a subject compares x to particular exemplars of A. However, exemplar models are not the only way to implement the instantiation principle. It is also possible to represent idiosyncratic or instance-level information in abstraction models, while simultaneously maintaining central tendencies or modal values. In the extreme, it is possible to construct a rich abstract representation that is informationally equivalent to storing all instances of a category (Barsalou, 1990). For example, in the abstraction model of Reitman and Bower (1973), information about all individual features, including idiosyncratic ones, and all combinations of features is represented. Neumann (1974) and Hayes-Roth and Hayes-Roth (1977) have proposed similar models.

It is important to note that we are not trying to distinguish between the broad classes of exemplar and abstraction models. On the contrary, our goal is simply to evaluate the instantiation principle, assuming that this principle could be implemented in many ways. Following Marr (1982), we assume that the instantiation principle lies at the computational level of theory, and that exemplar and abstract implementations of this principle lie at the algorithmic level (also see Anderson, 1991). In the studies to follow, we seek evidence for the instantiation principle, but we do not attempt to establish that one kind of model implements this principle better than another. Because many models of categorisation implement cognitive economy, rather than the instantiation principle, we believe that assessing this issue at a general level is an important and potentially informative pursuit.

It is also important to note that we are not claiming the instantiation principle to be a complete account of categorisation. In no way do we suggest that instantiation is sufficient to account for all categorisation phenomena. To the contrary, we believe that many other cognitive mechanisms must be involved as well. For example, a simple instantiation mechanism cannot account for the conceptualisation of entities never encountered. Thus, our goal in the following studies is simply to assess whether the instantiation principle should be included in the set of principles needed to account for human categorisation. Should the evidence support this principle, it would provide an important constraint on future accounts.

In three studies, we used typicality judgments to assess the instantiation principle. All categories, even those that are well-defined, exhibit typicality (Armstrong, Gleitman, & Gleitman, 1983; Barsalou, 1987). Some members of any category are judged reliably as better members than others. Moreover, typicality is arguably a better overall predictor of category performance than any other variable, predicting category verification time, exemplar production, category acquisition, and categorical reasoning. Interestingly, typicality gradients do not remain fixed across contexts, between people, or even within a single person over time (Barsalou, 1987, 1989, 1993). An exemplar that is typical in one context may be atypical in another; an exemplar typical for one person may be atypical for another; an exemplar typical for a person on one occasion may be atypical on another essentially similar occasion.

The instantiation principle has clear implications for judgments of typicality. If the instantiation principle holds, then subjects' typicality judgments for the member of a category should be influenced by detailed information about the member's instantiations. To see this, imagine evaluating the typicality of the subordinate category, *reptile*, within the superordinate category, *animals*.[1] According to the instantiation principle, subjects should use idiosyncratic knowledge about different kinds of reptiles to assess the typicality of *reptile*, not just general features true of most reptiles. Detailed information about specific exemplars, such as *snake*, *lizard*, and *alligator*, should enter into the representation of *reptile* and affect its typicality in *animal*.

In the following three studies, we asked subjects to judge the typicality of subordinate categories (e.g. *reptile*, *mammal*, *fish*, *insect*) in a superordinate category (e.g. *animal*). For each subordinate (e.g. *reptile*), we collected detailed information about its instantiations (e.g. *snake*, *lizard*, *alligator*). This information included the production frequency for each instantiation (e.g. how often subjects produce *snake* for *reptile*) and its typicality in the superordinate (e.g. how typical *snake* is of *animal*). If the instantiation principle is correct, then the distribution of instantiations for a subordinate, along with the typicalities of these instantiations, should predict the subordinate's typicality in the superordinate. If subjects sample information from a subordinate's instantiations to judge its typicality, then the distributional properties of these instantiations should predict its typicality.

Specifically, we predict, first, that the mean typicality of a subordinate's instantiations, weighted for production frequency, will predict the mean typicality judgments of the subordinate itself. For example, the average typicality of *reptile*'s instantiations in *animal* should predict the typicality of *reptile* itself in *animal*. Second, we predict that the variability of the typicality judgments for a subordinate's instantiations will predict the variability of the typicality judgments for the subordinate itself. (Assessing variability serves as a way of measuring the instability of typicality.) Third, we predict that the skew of the typicality judgments for a subordinate's instantiations will predict the skew of the typicality judgments for the subordinate itself (we only test this hypothesis in Study 3, where appropriate conditions exist for doing so). Finally, we predict that systematically dropping out the atypical instantiations of a subordinate should decrease the ability of the remaining instantiations to predict the subordinate's typicality. If all of the information in a distribution enters into a subordinate's typicality, even information from the least typical exemplars, then removing them should decrease prediction (we only test this hypothesis in Study 3, where we develop the richest data set).

Note that the instantiation principle would further predict that detailed exemplar information should enter into the representation of the superordinates such as *animal*, not just into the representation of subordinates such as *reptile*.

[1] Note that we use *subordinate* and *superordinate* here only in the sense of one category being a subordinate of a second, more superordinate, category. We are not using these terms to denote the subordinate and superordinate levels discussed in research on basic level categories (e.g. Rosch et al., 1976).

Although we assume that the instantiation principle applies to both the subordinate and superordinate category in a typicality judgment, the following studies only assess its role in subordinates. Should these categories reflect the instantiation principle, the superordinates may reflect it as well, although direct evidence would be necessary from future experiments. A related issue is whether the instantiations of subordinates are themselves instantiated. For example, when *lizard* instantiates *reptile*, do specific individuals or subcategories instantiate *lizard*, rather than representing this category with its general features? How far down does the instantiation process go? Again, we assume that the instantiation process could be more extensive than our present analysis assumes, extending both above and below the categories whose instantiation we measure.

A Model for Assessing the Instantiation Principle

To assess the role of instantiation in typicality, we have developed an exemplar-based model of instantiation (Heit & Barsalou, 1992). The primary virtue of this model is that it allows us to test the instantiation principle in a straightforward and conservative manner. Our goal is not to establish this model as a better account of the instantiation principle than other possible implementations of this principle. For example, we do not develop a competing abstraction model that also implements the instantiation principle. Instead, our goal in developing this exemplar model is simply to use it as a means for assessing the more general instantiation principle.

According to this model, people judge the typicality of a subordinate in a superordinate by performing three steps: first, a single instantiation of the subordinate category is retrieved (e.g. *lizard* is retrieved to instantiate *reptile*). Second, the instantiation's typicality in the superordinate is determined (e.g. the typicality of *lizard* in *animal* is judged as 3 on a 1–9 scale, where 1 is atypical and 9 is typical). Third, the instantiation's typicality is generalised to the subordinate (e.g. the typicality of *reptile* in *animal* is rated as 3).

The key sources of information that this model uses to predict performance are, first, a distribution of instantiations for the subordinate, and second, the typicality of these instantiations in the superordinate. The model uses no other information and has no free parameters. Again, this simple model is intended as a quantifiable means of assessing whether representations of subordinates include information about their full ranges of instantiations. If typicality judgments follow the instantiation principle, then the instantiation model should fit their data to a reasonably good extent. Clearly, there may be significant departures from the predictions of such a simple model, but these may well be accounted for by further developments of the model.

Equation 1 describes the model formally. T(A) is the typicality of subordinate category A, which has n instantiations or distinct category members, a_i. Equation 1 defines the probability of subordinate A receiving a typicality rating of j, where j is an integer on a predefined rating scale. In effect, the probability of A receiving a typicality rating of j is the sum of the probabilities that each of the subordinate category's instantiations, a_i, receives a typicality rating of j, weighted by the probability that each of the instantiations, a_i, is retrieved for A.

$$P(\mathrm{T}\,(A) = j) = \sum_{i=1}^{n} P(A \text{ is instantiated as } a_i)\; P(\mathrm{T}\,(a_i) = j) \tag{1}$$

It is important to note that $\mathrm{T}(A)$ and $\mathrm{T}(a_i)$ both refer to typicality in the same superordinate category. For example, $\mathrm{T}(A)$ might refer to the typicality of *reptile* in *animal*, and $\mathrm{T}(a_i)$ might refer to the typicality of *lizard* in *animal*. ($\mathrm{T}(a_i)$ would *never* refer to the typicality of *lizard* in *reptile*.)

Table 1 illustrates how the instantiation model might simulate the distribution of responses to the question, "How typical are mammals of animals?". Suppose that eight subjects are each asked to produce one instantiation of the subordinate

TABLE 1

Sample Application of the Instantiation Model

Instantiation	*Production Frequency*	*Typicality with Respect to Animal*
Human	3	9, 8, 9, 9, 1, 5
Bear	2	9, 6, 7, 9, 9, 9
Dog	1	9, 9, 9, 9, 9, 5
Kangaroo	1	4, 8, 9, 4, 1, 6
Whale	1	8, 7, 9, 7, 6, 9

Predicted Distribution for Mammal

Derived from Ratings on:	*Typicality with Respect to Animal*
Human	9, 8, 9, 9, 1, 5
Human	9, 8, 9, 9, 1, 5
Human	9, 8, 9, 9, 1, 5
Bear	9, 6, 7, 9, 9, 9
Bear	9, 6, 7, 9, 9, 9
Dog	9, 9, 9, 9, 9, 5
Kangaroo	4, 8, 9, 4, 1, 6
Whale	8, 7, 9, 7, 6, 9

Observed Distribution for Mammal

Derived from Ratings on:	*Typicality with Respect to Animal*
Mammal	9, 7, 9, 8, 2, 8

In this illustration, the predicted mean for *mammal* is 7.3 and the predicted standard deviation is 2.5. The observed mean for *mammal* is 7.2 and the observed standard deviation is 2.7.

mammal. As shown in Table 1, three subjects produce *human*, two produce *bear*, and one each produces *dog*, *kangaroo*, and *whale*. Next, a second group of six subjects rate the typicality of these instantiations in *animal* on a 1–9 scale (shown in the top section of Table 1, in the third column).

These two sources of data, production frequencies and typicality judgments, are used to predict the responses of 48 simulated subjects, as illustrated in the middle section of Table 1. (Forty-eight subjects are simulated, because eight subjects produced one exemplar each, and six ratings were obtained for each exemplar.) Each simulated subject first instantiates *mammal* according to the proportions obtained from the production task (e.g. *bear* has a 2/8 chance of instantiating *mammal*). Next, the simulated subject judges the typicality of the instantiation in *animal*, according to the instantiation's distribution of typicality ratings (e.g. *bear* has a 4/6 chance of being rated 9).

To evaluate the instantiation model, its predicted distribution of typicality ratings, derived from simulated subjects, is compared to an observed distribution of typicality ratings obtained from human subjects (illustrated in the bottom section of Table 1). For example, the predicted distribution for *mammal*, derived from the typicality of its instantiations, is compared to a distribution of typicality ratings obtained from subjects who directly rated the typicality of *mammal* in *animal.* To the extent that the predicted and observed distributions of ratings are similar for mean, variability, and skew, the instantiation principle receives support.

Overview

Each of three studies included two groups of subjects: production subjects and rating subjects. Production subjects provided a single instantiation for each subordinate (e.g. name the first *mammal* that comes to mind). The relative frequencies of these instantiations were used to estimate the value of $P(A$ is instantiated as $a_i)$ in Equation 1 (e.g. the probability that *goat* is retrieved for *mammal*). Rating subjects then judged the typicality of these instantiations with respect to the superordinate category (e.g. judge the typicality of *goat* in *animal*). These judgments were used to estimate the values of $P(\mathrm{T}(a_i)=j)$ in Equation 1 (e.g. the probability that *goat* receives a typicality rating of 5 for *animal*). The rating subjects also judged the typicality of the subordinate categories themselves, thereby providing the values of $P(\mathrm{T}(A)=j)$ in Equation 1 (e.g. the typicality of *mammal* in *animal*).

Studies 1 and 2 were similar, with Study 1 assessing the instantiation principle for seven subordinates of *animal*, and Study 2 assessing the instantiation principle for nine subordinates of *food.* Study 3 was more elaborate, assessing the role of the instantiation principle in complex categories of animals (e.g. *small fish*) as well as in simple categories (e.g. *fish*). In this study, the complex categories were 42 of the 63 subordinates (e.g. *dangerous insect*) and 2 of the 3 superordinates (e.g. *dangerous animal*). Thus, Study 3 assessed the instantiation principle for a more challenging variety of categories than in Studies 1 and 2.

STUDY 1

Method

Subjects. The production and rating groups each consisted of 20 Stanford University undergraduates, recruited in residence halls. The production subjects were all run before the rating subjects.

Materials. For the production group, the category cues were seven subordinates of *animal*: *amphibian*, *bird*, *fish*, *insect*, *mammal*, *micro-organism*, and *reptile.* The production subjects produced a total of 63 unique instantiations in response to these seven cues. For the rating group, the rated categories were the 63 instantiations produced by the production group, plus the 7 subordinate categories. Each instantiation appeared only once as a rating stimulus, regardless of the number of production subjects who mentioned it.

Procedure. Production subjects performed the experiment individually. Each subject was instructed to name the first instance that came to mind for each category cue. The seven cues were read to the subject one at a time in a random order. The experimenter wrote down the subject's response, then proceeded to the next category. Subjects were not prevented from making responses that were scientifically incorrect (e.g. producing *duck* as an instantiation of *amphibian*).

Rating subjects were instructed to rate the 7 subordinate categories and the 63 instances on typicality with respect to *animal*, using a 1–10 scale, for which higher numbers meant greater typicality. For example, subjects were asked, ‘‘How typical is a *dog* for the category *animal*?’’ and ‘‘How typical is a *mammal* for the category *animal*?’. Note that subjects never rated the instantiations on their typicality in the subordinate categories. For example, subjects never evaluated the typicality of *dog* in *mammal*. Subjects performed the rating task at their own pace, normally taking 5 to 10 minutes.

Results and Discussion

Figures 1 and 2 display, on their abscissas, the observed means and standard deviations, respectively, of the typicality ratings for the seven subordinates. Each data point is shown as the first letter or two of the category (e.g. ‘‘Ma’’ for *mammal*). As Fig. 1 illustrates, *mammal* was rated most typical, and *micro-organism* was rated least typical. Figure 2 indicates that the ratings were most stable for *micro-organism* and least stable for *fish*, *amphibian*, and *insect.*

To test the instantiation principle, the instantiations from the production subjects and the typicality judgments from the rating subjects were used to

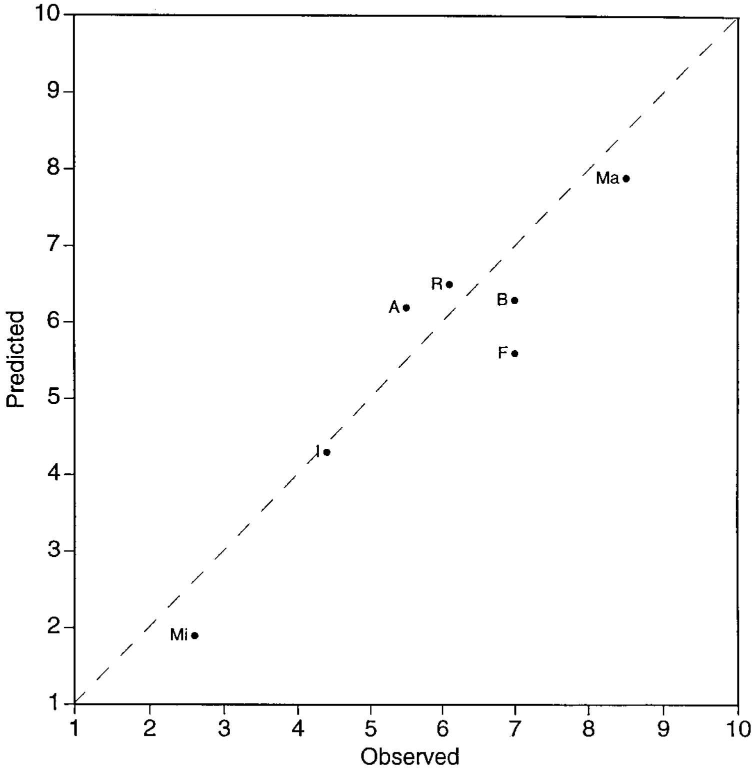

FIG. 1. Observed and predicted means of typicality ratings for *animal*, Study 1.

compute the predictions of the instantiation model. Appendix B, which presents all of the instantiations for Study 3, provides a sense of the instantiations produced in Study 1 (i.e. the subordinates for Study 3 included those for Study 1).

Of primary concern is whether the seven predicted distributions, obtained by application of the instantiation model, corresponded to the seven observed distributions of the typicality ratings. The observed distribution for a subordinate was simply the 20 ratings that rating subjects made for it directly. The predicted distribution resulted from a simulation of the instantiation process, with each simulated subject run as follows. First, a subordinate was instantiated according to the production frequencies of its instantiations. For example, 5 of the 20 production subjects instantiated *mammal* with *human*. As a result, *human* was used 25% of the time on this step to predict the typicality of *mammal*. Second, a typicality rating for the instantiation was chosen randomly from the 20 typicality judgments that the rating subjects made for it. In the example for *human*, one of the 20 ratings that subjects made for *human* was chosen to represent the typicality of *mammal* in *animal*.

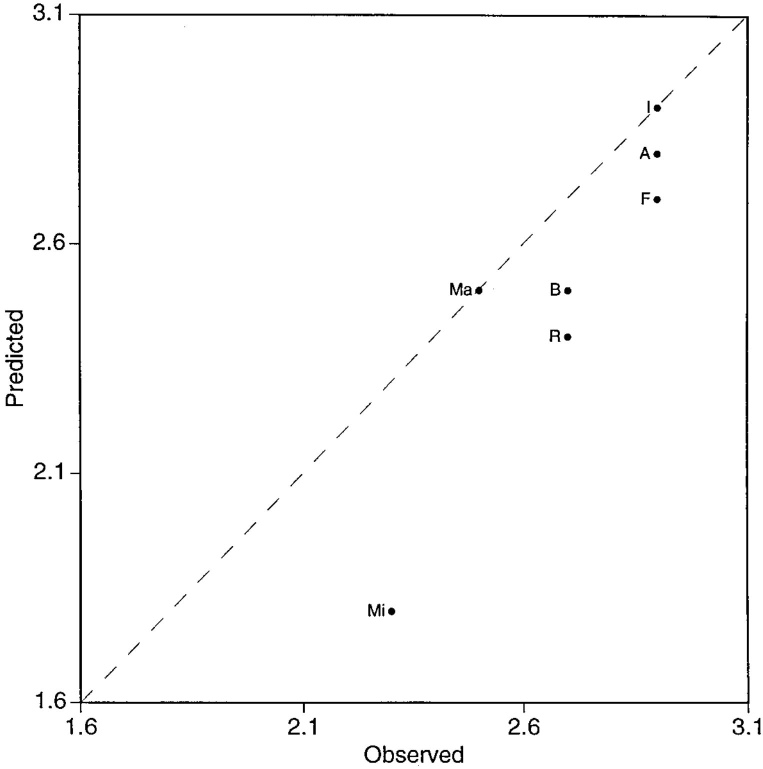

FIG. 2. Observed and predicted standard deviations of typicality ratings for *animal*, Study 1.

Overall, a total of 400 subjects were simulated for each subordinate in 20 cycles of 20 simulated subjects each. This procedure ensures that (1) the relative frequency of different instantiations is maintained in the simulated data, and (2) all rating data enter into the simulations. To satisfy the first constraint, the simulation exhaustively sampled the same 20 instantiations of a subordinate in each cycle, thereby maintaining the same relative frequency of instantiations across all cycles. Thus, *human* instantiated *animal* on exactly 25% of the simulations in every cycle. To satisfy the second constraint, the simulation sampled the ratings for an instantiation without replacement until the entire set of ratings had been sampled. If further ratings were needed for the instantiation, the same ratings were sampled again without replacement. Thus, all 20 ratings for a given instantiation were used at least once. For example, all 20 ratings for *horse* were used once and only once, because *horse* had a production frequency of 1 and was included once in each of the 20 cycles. In contrast, all 20 ratings for *human* were used 5 times each, because *human* had a production frequency of 5 and was included 5 times in each of the 20 cycles.

Figures 1 and 2 display, on their ordinates, the predicted means and standard deviations, respectively, of the typicality ratings for the seven subordinates. The correspondence between the observed and the predicted means is excellent; $r = 0.93$, $P < 0.01$. Likewise, the correlation between the observed and predicted standard deviations is excellent; $r = 0.92$, $P < 0.01$.[2] (All P values for model fitting are reported for one-tailed hypotheses.) These outcomes are not attributable to extreme values unduly affecting the correlation coefficients; the results are similar after transforming the means and standard deviations to rank orders. For the means, $r_s = 0.79$, $P < 0.05$, and for the standard deviations, $r_s = 0.88$, $P < 0.05$.

These high correlations support the instantiation principle. As reflected in the good fits of the instantiation model, detailed information about the instantiations of subordinates predicts the typicality of these subordinates in a superordinate category.

A More Conservative Analysis of the Standard Deviations. Problematically, the correlations for the standard deviations could reflect, in part, a property of the rating scale. Note that the most unstable categories, *fish*, *amphibian*, and *insect* in Fig. 2, each have a mean typicality in Fig. 1 near the scale's midpoint. In contrast, *micro-organism* and *mammal* obtained more stable ratings and had mean typicality ratings near the endpoints of the scale. Thus, the higher standard deviations near the middle of the scale could reflect more available response categories near the scale's midpoint than near an endpoint. Because the observed and predicted means correlated so highly, the predicted and observed means for a given category tend to lie in the same region of the scale. As a result, the same availability of response categories for both means could have artificially enhanced the correlation of their standard deviations.

To control for this possible artifact, regression was used to remove variability associated with the rating scale. The dependent variable, observed standard deviation, was regressed onto the distance between the observed mean and the nearer endpoint of the response scale (1 or 10). The correlation coefficient for this regression analysis was 0.80, which supports the observation that standard deviations tended to decrease towards the ends of the scale. The simulated data exhibited the same pattern. When the predicted standard deviations were regressed onto the distance between the predicted means and the nearer endpoint of the scale, a correlation of 0.79 obtained.

To remove these effects from the fits of the model for standard deviations, the residuals from the two regressions were analysed. In the residuals, variability at

[2] Our rationale for evaluating the instantiation model with the r statistic is that it allows us to regress this measure of fit onto multiple sources of prediction. In later analyses, we will partial out various factors to assess their contributions to the model's ability to fit the data. Study 3 will also report RMSEs for fit to assess additional issues of interest.

the middle of the scale should be comparable to variability near the endpoints. The correlation between the residuals for the observed and predicted standard deviations was still positive, $r = 0.63$, $P = 0.06$. After transforming the residuals to rank orders, $r_s = 0.36$, n.s.

To summarise, a conservative analysis can remove variability that results from extra response categories near the middle of the scale. The fit of the instantiation model for standard deviations was somewhat worse after this procedure. However, this analysis may also remove differences among concepts that are truly reflective of an instantiation process, to the extent that the availability of response categories and the instantiation process are statistically related. Studies 2 and 3 had a larger number of data points and clearly demonstrate the independent role of the instantiation process.

STUDY 2

To provide generality, Study 2 was an attempt to replicate Study 1 with different categories. Production subjects produced instantiations for nine subordinates of *food*, and rating subjects judged the typicality of these instantiations in the superordinate.

Method

Subjects. The production group consisted of 40 University of Michigan undergraduates, recruited in public places in Ann Arbor. The rating group consisted of 40 Michigan undergraduates, who participated as part of a course requirement.

Materials. For the production group, the category cues were nine subordinates of *food*: *beverage*, *dairy product*, *dessert*, *fish*, *fruit*, *meat*, *poultry*, *seasoning*, and *vegetable*. For the rating group, the rated categories were the nine subordinates plus the 88 unique instantiations that the production group produced. Each subject received a page containing these food categories in one of four random orders distributed evenly among subjects.

Procedure. The procedure was like Study 1, except that rating subjects judged the categories on a scale from 1–9 with respect to their typicality in *food*.

Results and Discussion

Figures 3 and 4 display, on their abscissas, the observed means and standard deviations, respectively, of the typicality ratings for the nine subordinates. Each data point is shown as the first letter or two of the category (e.g. ‘‘Be’’ for *beverage*). As Fig. 3 illustrates, fruit and *meat* were rated most typical, and

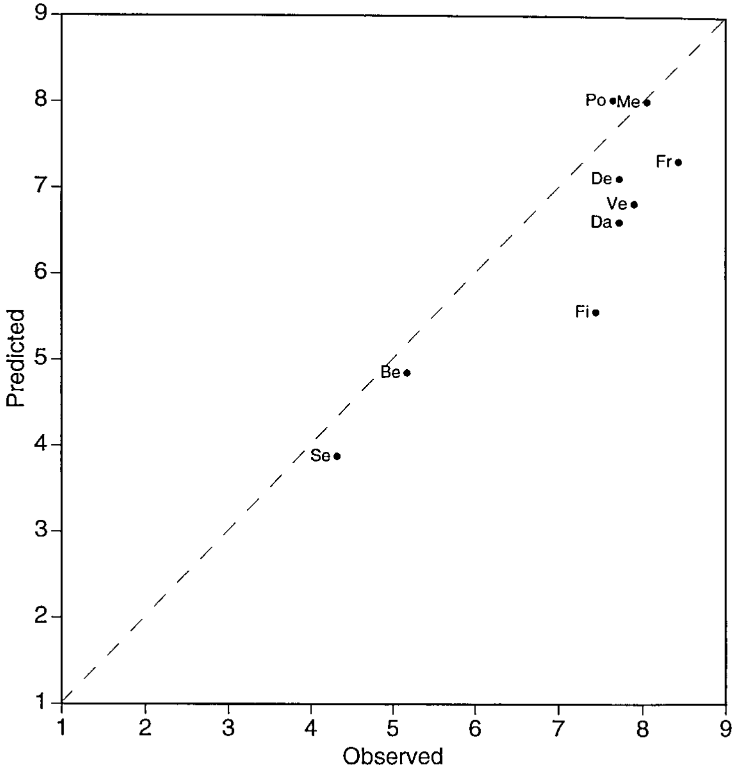

FIG. 3. Observed and predicted means of typicality ratings for *food*, Study 2.

seasoning and *beverage* were rated least typical. Figure 4 indicates that ratings were most stable for *fruit* and least stable for *seasoning* and *beverage*.

Again, of primary interest was how well the predicted typicality distributions of typicality ratings for the subordinates fit the observed distributions. Using the same simulation process described for Study 1, the predicted distribution for each of the nine subordinates was obtained by exhaustively simulating the 1600 possible combinations of instantiations and typicality ratings derived from the production and rating subjects. Appendix A presents the instantiations that the production subjects generated.

Figures 3 and 4 display, on their ordinates, the predicted means and standard deviations, respectively, of the typicality ratings for the nine subordinates. The correspondence between observed and predicted means is excellent, $r = 0.89$, $P < 0.01$. (Similarly, for rank-transformed data, $r_s = 0.70$, $P < 0.01$). For the standard deviations, $r = 0.64$, $P < 0.05$. (Likewise, for rank-transformed data, $r_s = 0.65$, $P < 0.05$.) These results are similar to Study 1, and again support the instantiation principle.

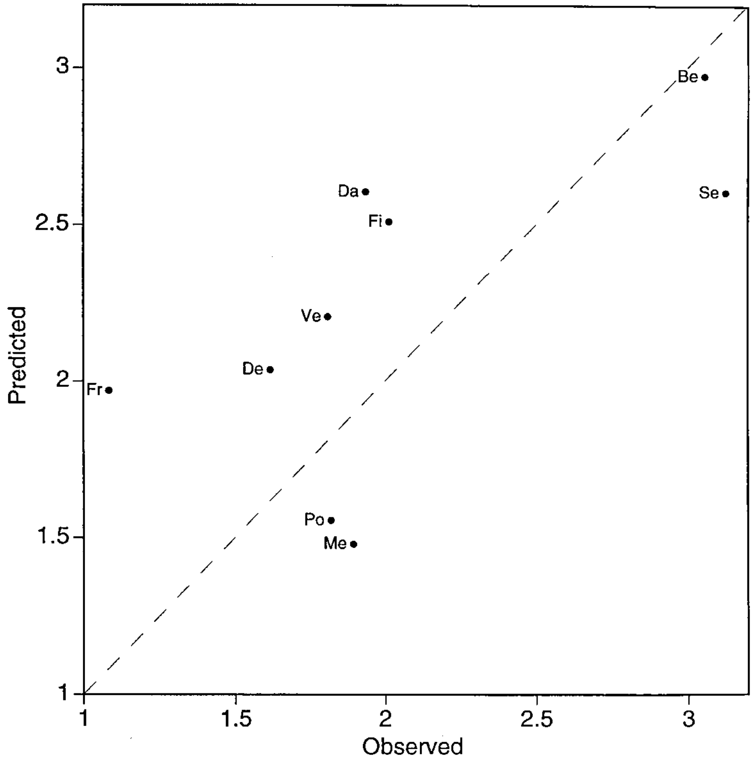

FIG. 4. Observed and predicted standard deviations of typicality ratings for *food*, Study 2.

Inspection of Figs. 3 and 4 indicates that, as in Study 1, there was extra variability associated with mean ratings near the middle of the rating scale. For example, *seasoning* and *beverage* received mean typicality ratings near 5, and these categories had the greatest variability. Using the same procedure as in Study 1, the standard deviations were regressed onto the distance between the means and the nearer endpoint of the scale, 1 or 9. These analyses showed that the observed standard deviations were predictable from this distance measure, $r = 0.95$, as were the predicted standard deviations, $r = 0.94$. Using the residuals from these two analyses, the correlation between the observed and predicted standard deviations was 0.65, $P < 0.05$. Using rank-transformed data, $r_s = 0.60$, $P < 0.05$. Thus, removing variability associated with the rating scale did not have a detrimental effect on the ability of the instantiation model to fit the standard deviations. Even after this conservative procedure, there was a significant correlation between the observed and predicted standard deviations.

STUDY 3

This study had three purposes: First, we examined the ability of the instantiation model to predict a wider range of judgments, including the typicality of complex categories such as *dangerous micro-organism.* Second, Study 3 included a much larger number of judgments overall than our previous studies, enabling a decisive analysis of standard deviations, as well as an additional analysis of skewness. Third, we investigated different versions of the instantiation model that made different assumptions about how people might instantiate categories. For example, we evaluated a new version of the model that only instantiated subordinates with high-frequency instantiations, such as the top 50% of the distribution produced by production subjects. If the instantiation principle is correct, then dropping the atypical instantiations from the simulations should worsen prediction significantly. On the other hand, if the instantiation principle is incorrect, and if subjects represent subordinates only with the properties typically true of most exemplars, then dropping atypical exemplars from the simulations should not worsen prediction. Because the typical exemplars embody the properties true of most exemplars, the typical exemplars would be sufficient for predicting typicality. Indeed, one could argue that dropping atypical exemplars should improve prediction by removing spurious sources of prediction from the simulations.

Study 3 examined 21 subordinates of *animal.* These subordinates included the seven subordinates from Study 1 (e.g. *mammal*, *reptile*), as well as 14 complex categories formed by crossing these seven simple categories with two modifiers, *small* and *dangerous* (e.g. *small mammal*, *dangerous reptile*). Another change from the previous studies was that we collected typicality judgments for three superordinates, not just one. Like Study 1, some subjects rated the 21 subordinates for typicality in *animal*; however, a second group rated these subordinates for typicality in *small animal*, and a third group rated them for typicality in *dangerous animal.* Thus, we assessed the ability of the instantiation model to handle complex categories at two levels: first, at the level of subordinates (e.g. *reptile*, *small reptile*, *dangerous reptile*), and second, at the level of superordinates (i.e. *animal*, *small animal*, *dangerous animal*).

We expected that the typicality of the subordinates would shift dramatically across the three superordinates. For example, the typicality of *insect*, *reptile*, and *mammal* could change considerably, depending on whether they are being judged relative to *animal*, *small animal*, or *dangerous animal.* Because typicality in superordinates varies dramatically across points of view (Barsalou, 1987, 1989), typicality appeared likely to vary across a similar type of manipulation here. Of interest was whether the instantiation process can generate accurate predictions as typicality shifts. For example, can the instantiations of *reptile* predict its shifting typicality in *animal*, *small animal*, and *dangerous animal*? Similarly, can the instantiations of *dangerous reptile* predict its shifting typicality in these same superordinates? Should the instantiation model succeed at predicting shifting typicality gradients, it would implicate the instantiation principle in a broader range of conceptual processing.

Method

Subjects. The subjects for the production task were 120 University of Michigan undergraduates who performed the task in public places for no compensation. These subjects were assigned randomly to one of three conditions, unmodified, small, or dangerous, with 40 subjects in each. The subjects in the rating task were 96 University of Michigan undergraduates who received course credit for participation. Subjects were assigned randomly to one of three conditions, unmodified, small, or dangerous, with 32 subjects in each.

Materials. In the unmodified condition of the production task, the category cues were the seven subordinates from Study 1: *amphibian*, *bird*, *fish*, *insect*, *mammal*, *micro-organism*, and *reptile.* In the small condition of the production task, the category cues were these seven categories modified by the adjective *small*. In the dangerous condition of the production task, the category cues were the seven simple categories modified by the adjective *dangerous.*

For the rating task, the stimuli were the 21 animal categories used as category cues in the production task (e.g. *mammal*, *small amphibian*, *dangerous reptile*) and the 186 unique instantiations obtained as responses. Because of some redundancy, the total number of rated categories was 205 (e.g. the category *fish* was both a cue in the production task and a response from a subject to another category). Each subject received a booklet containing these animal names in one of four random orders distributed evenly among subjects.

Procedure. The procedure was like Study 2, except for the following changes: Production subjects produced instantiations for only 7 of the possible 21 subordinates (unmodified, small, or dangerous). Each subject only received seven category cues to prevent carryover effects from one category to another (e.g. instantiating *small bird* might affect the subsequent instantiation of *bird*).

For the rating task, subjects in the unmodified condition were instructed to judge typicality with respect to *animal*; subjects in the small condition rated typicality with respect to *small animal*; subjects in the dangerous condition rated typicality with respect to *dangerous animal.* Subjects only rated typicality in one superordinate. For example, some subjects rated the typicality of *trout* in *animal*, and other subjects rated the typicality of *trout* in *small animal.* Subjects performed this task at their own pace, normally taking 15 to 20 minutes.

Results and Discussion

Before assessing the fit of the instantiation model, we describe the shifts in to the instantiation model's performance will be how well it handles these shifts.

The production task yielded 40 instantiations for each of the 21 subordinates, with multiple subjects often producing the same instantiation. Appendix B presents the distribution of instantiations for each subordinate. The manipulation of the adjective modifier had a substantial effect on production. For example, subjects never produced *hummingbird* or *hawk* as instantiation of *bird*, yet they produced *hummingbird* most frequently for *small bird*, and they produced *hawk* most frequently for *dangerous bird.* Casual perusal of Appendix B indicates that the adjective modifiers produced major shifts in the production of instantiations for each subordinate.

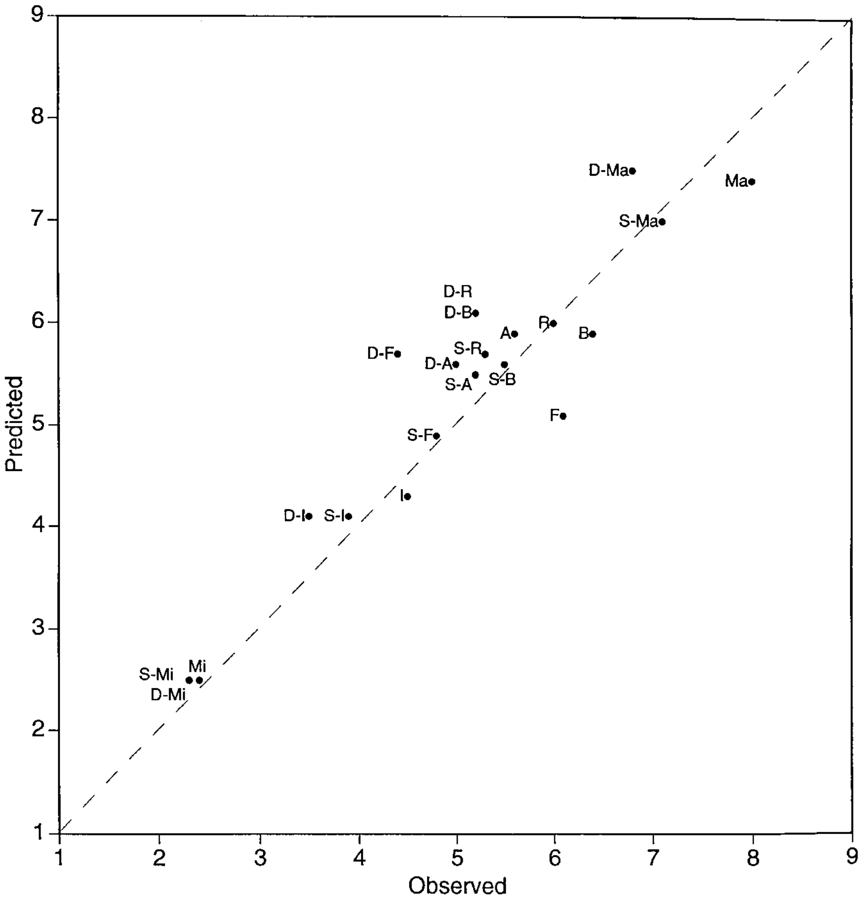

FIG. 5. Observed and predicted means of typicality ratings for *animal*, Study 3.

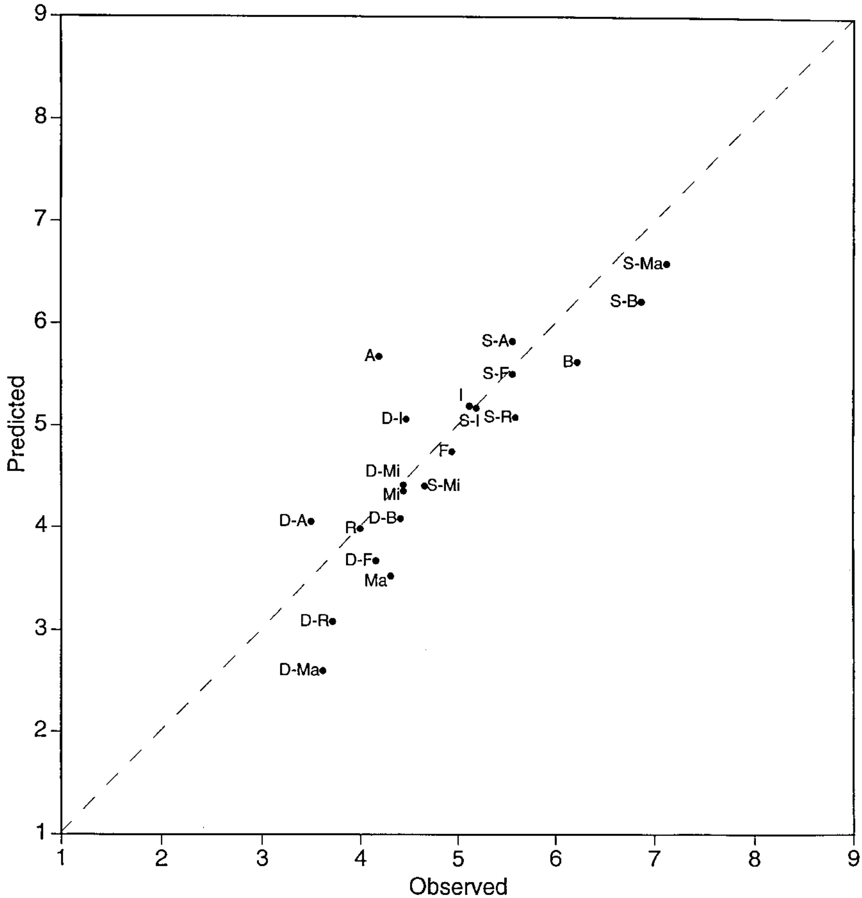

FIG. 6. Observed and predicted means of typicality ratings for *small animal*, Study 3.

Figures 5, 6, and 7 display, on their abscissas, the observed means of the typicality ratings for the 21 subordinates in *animal*, *small animal*, and *dangerous animal*, respectively. Figures 8, 9, and 10 display, on their abscissas, the observed standard deviations of the typicality ratings for the 21 subordinates in *animal*, *small animal*, and *dangerous animal*, respectively. Modified subordinates are indicated with the prefix ''S'' for *small* and ''D'' for *dangerous* (e.g. ''S–F'' for *small fish*, ''D–R'' for *dangerous reptile*).

Comparison of Figs. 5, 6, and 7 indicates that the mean typicality judgments for the 21 subordinates shifted considerably across the three modifier conditions. In the unmodified condition, *mammal* was rated as typical, whereas *micro-organism* was rated as atypical. Interestingly, the adjective modifiers had little

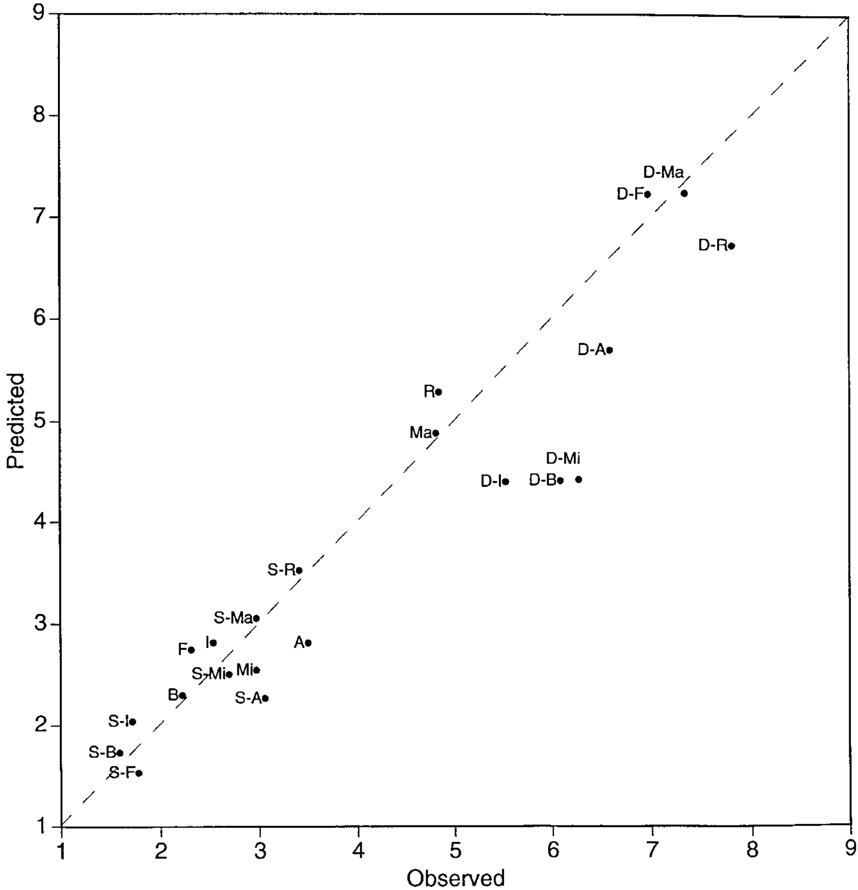

FIG. 7. Observed and predicted means of typicality ratings for *dangerous animal*, Study 3.

effect in the unmodified condition. For example, the mean ratings for *insect*, *small insect*, and *dangerous insect* were quite similar, as were the three sets of ratings for every other head noun. In the small condition, the target categories modified by *small* were rated most typical, whereas the categories modified by *dangerous* were rated least typical. A similar shift occurred in the dangerous condition, where categories modified by *dangerous* were rated most typical, and categories modified by *small* were rated least typical. Comparison of Figs. 8, 9, and 10 indicates that the standard deviations also shifted somewhat with context, although not nearly as much as the means.

Applying the Instantiation Model. Of primary interest was whether the simulated distributions of typicality judgments for the 21 target categories

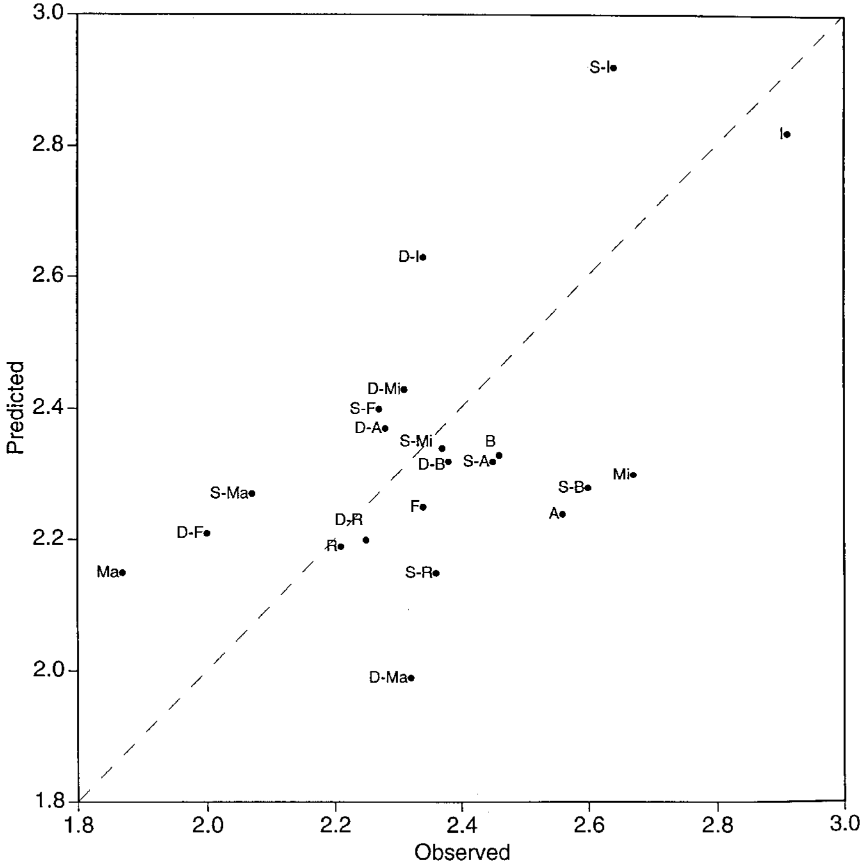

FIG. 8. Observed and predicted standard deviations of typicality ratings for *animal*, Study 3.

predicted the observed typicality distributions. Three distributions were simulated for each of the 21 subordinate categories, one for each superordinate, *animal*, *small animal*, and *dangerous animal.* For a given subordinate, the same set of instantiations from the production task (Appendix B) was used to construct the simulated distribution for each superordinate. For example, the same set of instantiations for *small fish* was used to generate the predicted distributions for its typicality in *animal*, *small animal*, and *dangerous animal.* Thus, the differences between the three predicted distributions only reflected the different typicality ratings obtained for a common set of instantiations. The fixed distribution of instantiations for a given subordinate reflects a strong assumption that the initial step of instantiating a subordinate is relatively encapsulated from

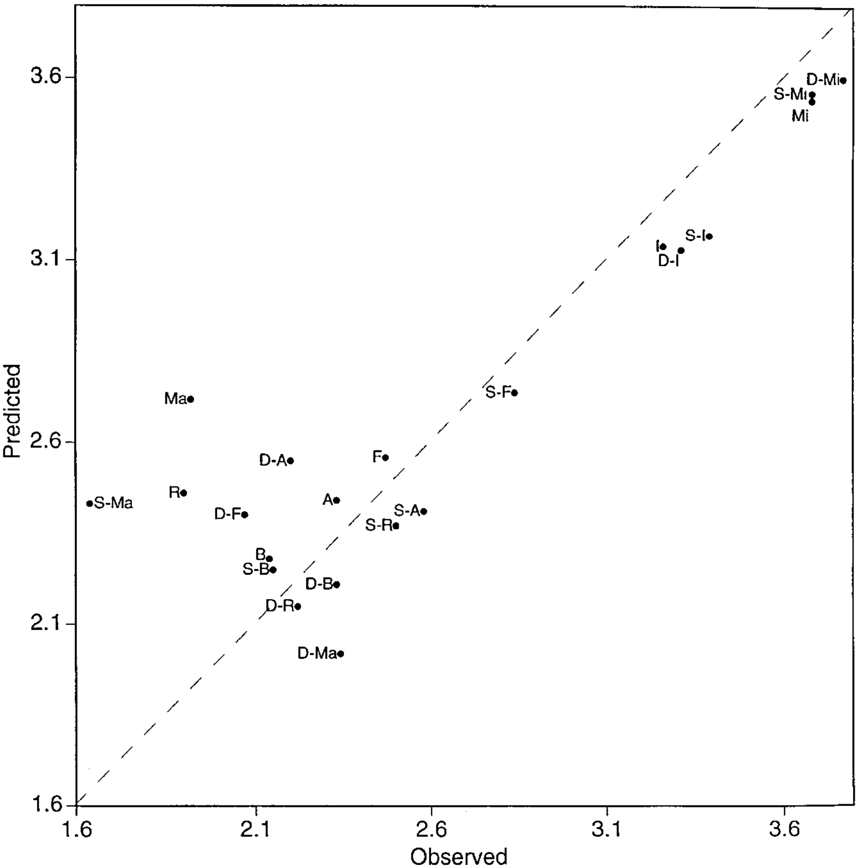

FIG. 9. Observed and predicted standard deviations of typicality ratings for *small animal*, Study 3.

the subsequent step of evaluating its typicality, which, in contrast, is highly context-dependent. To the extent that the instantiation model fits the data, this assumption is justified.

Figures 5, 6, and 7 display, on their ordinates, the predicted means of the typicality ratings for the seven subordinates in *animal*, *small animal*, and *dangerous animal*, respectively. As Fig. 5 illustrates, the instantiation model provides an excellent account of the observed means for typicality in animal, with the predicted means correlating 0.93 with the observed means. As Figs. 6 and 7 likewise show, the instantiation model also provides an excellent account of the means for *small animal* and *dangerous animal*, $r = 0.85$ and $r = 0.95$, respectively (all three Ps < 0.001). The results are similar after the data are

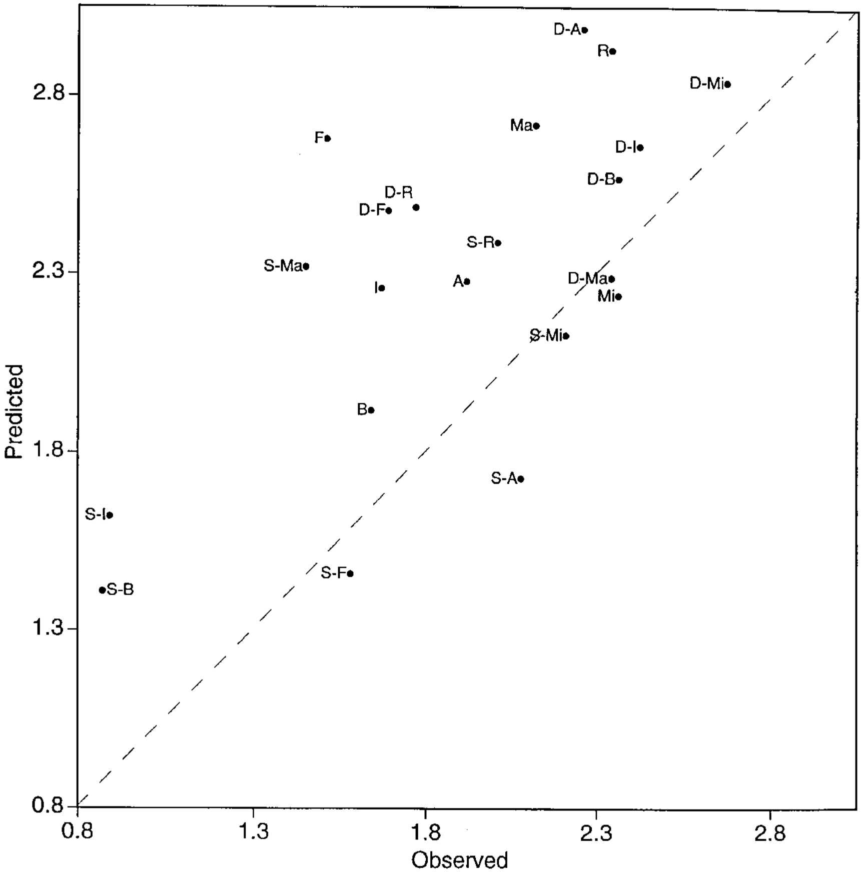

FIG. 10. Observed and predicted standard deviations of typicality ratings for *dangerous animal*, Study 3.

transformed into rank orders; the three respective Spearman correlation coefficients are 0.84, 0.85, and 0.92 (all three Ps < 0.001). These results suggest that the instantiation principle is just as central to the processing of complex categories as it is to the processing of simple categories.

Figures 8, 9, and 10 display, on their ordinates, the predicted standard divisions of the typicality ratings for the seven subordinates in *animal*, *small animal*, and *dangerous animal*, respectively. As these figures illustrate, the instantiation model accounts for a significant amount of the variation in the standard deviations. The predicted standard deviations correlated 0.56 ($P < 0.01$)

with the observed standard deviations for *animal*, 0.88 ($P < 0.001$) with the observed standard deviations for *small animal*, and 0.66 ($P < 0.001$) with the observed standard deviations for *dangerous animal*. For rank-transformed data, the respective correlations are $r_s = 0.44$, $P < 0.05$; $r_s = 0.63$, $P < 0.01$; $r_s = 0.54$, $P < 0.01$.

As in Studies 1 and 2, regression was used to remove variability due to the number of available response categories across the rating scale. All 63 possible subordinate–superordinate pairs were included in the analysis. When the distance of the observed means from the nearest endpoint was regressed onto the observed standard deviations, the correlation was 0.63, again suggesting that standard deviations tended to be higher near the midpoint of the scale. Similarly, when distance of the predicted means from the nearest endpoint was regressed onto the predicted standard divisions, the correlation was 0.64. Next, the residuals from the regression were analysed to assess whether the instantiation model accounted for significant variability in standard divisions that could not be attributed to the rating scale. The correlation was 0.73, $P < 0.001$. (For rank-transformed data, $r_s = 0.76$, $P < 0.001$.) These results indicate that the instantiation model accounts for substantial variability in the observed standard deviations, when possible confounding factors have been eliminated.

The ability of the instantiation model to predict standard deviations in Studies 2 and 3 suggests that it is reasonable to interpret the moderately high but not statistically significant correlation in Study 1 as also supporting the instantiation principle. In Study 1, with a smaller number of independent data points, the instantiation model's ability to predict standard deviations did not quite reach statistical significance after removing the possible influence of the rating scale. In contrast, Study 3 showed a correlation of 0.73 across 63 items, which is comparable to a correlation of 0.63 across 7 items in Study 1. Given the significant 0.65 correlation in Study 2, we are inclined to conclude that the instantiation model has done a good job of predicting standard deviations across all three studies.

To further compare the observed and predicted distributions, we assessed their skews. If, for example, the instantiations of a subordinate include a large number of typical exemplars and a small number of atypical exemplars (negative skew), then the instantiation model would predict that the distribution of ratings for the subordinate should be skewed in the same direction.

Unlike the standard deviations, the skew measures were not confounded with the mean ratings. For the 63 observed distributions, the correlation between mean and skew was only 0.11, and for the 63 predicted distributions, the correlation was 0.14. The critical finding is that the correlation between the skew of the observed distributions and the skew of the predicted distributions was 0.87, $P < 0.001$. (For rank-transformed data, $r_s = 0.84$, $P < 0.001$.) Not only were the observed and predicted distributions highly similar in their means and standard deviations, they were also highly similar in their skews.[3]

In summary, the instantiation model accounted for the typicality of complex categories at two taxonomic levels: as typicality gradients shifted across *animal*, *small animal*, and *dangerous animal*, the instantiation model did a good job of accounting for the typicality of both unmodified and complex subordinates. Furthermore, the instantiation model captured detailed properties of the observed distributions, simulating standard deviations and skews as well as means.

Alternate Versions of the Instantiation Model. The standard instantiation model assumes that the instantiation process during typicality judgments potentially reflects all instantiations produced by the production subjects. For example, the instantiation model assumes that *amphibian* has a 33% chance of being instantiated with *frog* and a 3% chance of being instantiated with *alligator*, because production subjects produced *frog* and *alligator* 33% and 3% of the time, respectively. This assumption follows from the instantiation principle, which states that the representation of a category reflects its full range of instantiations, weighted by their frequency. The success of the standard model in fitting the results of three studies provides considerable support for this assumption.

It is possible, however, that the representation of a category only contains features generally true of its members, as in classical and standard prototype models (e.g. Barsalou & Hale, 1993; Hampton, 1993; Smith & Medin, 1981). Properties true of only a single member, or of a few members, are not included, nor are properties true of atypical exemplars. On this view, a category's representation is most likely to contain only those features that are generally true of typical members. Thus, the representation of *mammal* might generally reflect knowledge about *dog*, which is a typical member of *mammal*, but not knowledge about *whale*, which is atypical. Whereas the representation might include the features *furry* and *walks*, it might not include the features *blow spout* and *swims*.

To investigate this possibility, we developed additional versions of the instantiation model. In these versions, we systematically dropped atypical exemplars from the simulations. Of interest was the effect that this had on the model's ability to predict the observed distributions. If the instantiation principle

[3] As in Study 3, the skews of the predicted distributions in Studies 1 and 2 correlated highly with the skews of the observed distributions. In Studies 1 and 2, however, the skews and means were largely confounded, with subordinates having high means tending to have negative skews. Similar to the standard deviations, the properties of the rating scale may have artificially enhanced the correlations between the predicted and observed skews. After regression was used to remove variability due to the rating scale, the instantiation model failed to predict any of the residual variance. Because this confound did not appear in Study 3, the regression analysis did not impair the ability of the instantiation model to predict skew.

is correct and all category information is potentially used in typicality judgments, then accuracy should decrease as atypical exemplars are removed from the simulations. On the other hand, if the instantiation principle is incorrect and only the most typical features of a category are used in typicality judgments, then accuracy should not be impaired by removing atypical exemplars, and it might even improve. If subjects do not use information from atypical exemplars, then removing this information could improve prediction by removing spurious sources of prediction from the model.

To see how these alternative simulations worked, consider the productions for *amphibian* in Appendix B. For the simulation that dropped the lowest 25% of the distribution, we excluded the following instantiations: *snake*, *alligator*, *crocodile*, *duck*, *fish*, *salmon*, *spider*. The remaining instantiations (*frog*, *lizard*, *salamander*, *turtle*, *toad*) constituted about 75% of the distribution.

We ran four new simulations that dropped 25%, 50%, 75%, or 90% of the lowest-frequency instantiations cumulatively, leaving 75%, 50%, 25%, or 10% of the highest-frequency instantiations. (The original simulations, in effect, dropped 0% of the least-frequent instantiations.) The additional simulations required a stochastic element, because there were occasional ties between instantiations having the same production frequency. For example, *bee* and *fly* were each produced by 15% of the subjects for *insect*. Thus, when the simulation dropped the bottom 90% of the instantiations, either *bee* or *fly* was selected randomly to remain. We ran 20 different random versions of each simulation to ensure that the results were not unduly influenced by chance. Thus, the following analyses indicate the average performance across 20 replications.

Table 2 presents the results of these simulations. As in previous analyses, we evaluated the simulations in terms of how well they predicted the mean, standard deviation, and skew of the 63 subordinates. Also, as in previous analyses, the primary measure of goodness-of-fit was the correlation coefficient. However, we also computed the root mean squared error (RMSE) to enable comparison between models. Although the absolute magnitude of the RMSE is somewhat difficult to evaluate, it is especially sensitive to the differential ability of different models to predict observed data.

TABLE 2

Evaluation of Differing Assumptions about Dropping Low-frequency Instantiations

	Mean		*Standard Dev.*		*Skewness*	
% Dropped	*r*	*RMSE*	*r*	*RMSE*	*r*	*RMSE*
0%	0.92	0.63	0.76	0.40	0.87	0.53
25%	0.92	0.66	0.75	0.40	0.86	0.63
50%	0.90	0.74	0.72	0.41	0.84	0.77
75%	0.86	0.94	0.69	0.45	0.75	0.88
90%	0.86	0.99	0.78	0.41	0.75	0.91

As Table 2 illustrates, the ability of the instantiation model to fit means and skews deteriorated as increasing numbers of low-frequency instantiations were dropped. Both the correlations and the RMSEs exhibited this effect, although the more sensitive RMSEs exhibited it more strongly. This finding indicates that subjects were not excluding information about atypical exemplars in their judgments. If they had, prediction should have improved or at least remained the same as atypical exemplars were dropped from the simulations. The fact that prediction deteriorated instead supports the instantiation principle. Consistent with previous results, subjects appear to be using all of the potentially available information in representing subordinates.

Although these simulations clearly implicate the use of atypical exemplars in subjects' judgments, we were surprised that massive deletions of exemplars did not cause prediction to deteriorate more than it did. For example, the simulations that kept only the 10% most frequent instantiations still performed relatively well, producing a correlation of 0.86 with the observed means. We attribute this finding to a redundancy of information across the instantiations of a given subordinate, namely, the different instantiations generally exhibited the same typicality in the superordinate. For example, all insects were viewed as relatively atypical of *animal.* Thus, even if low-frequency instantiations such as *beetle* and *grasshopper* were dropped from the simulation, higher-frequency instantiations such as *bee* and *fly* redundantly carried the same typicality information.

Another surprising finding was that the predictions for the standard deviations did not deteriorate as low-frequency instantiations were dropped. This finding, too, may be explained in part by redundancy within the instantiations for a given subordinate. Apparently, the variability of a subordinate's instantiations is relatively constant across levels of production frequency, such that removing atypical exemplars leaves the same variability within the ratings of typical exemplars. One way to interpret this result is that the instability of typical exemplars is roughly the same as the instability of atypical exemplars, a finding that we have observed in other work (Barsalou & Sewell, in prep.).

Another important factor may be the presence of an upper limit or ceiling on how well the instantiation model, or any model, can predict standard deviations. Inspection of the figures showing standard deviations (Figs. 2, 4, 8, 9, and 10) indicates that the standard deviations did not vary over a large range. In effect, this low range of variability may limit any model in predicting standard deviations. Because there is little in the way of variability to predict, small amounts of noise in measurement can undermine a model's ability to capture systematic variance. Similarly, dropping out exemplars may have little effect, because there remains relatively little variance to predict.

The relatively small RMSEs for standard deviations in Table 2 strongly support this conclusion. In terms of absolute error, the model is doing an excellent job of predicting the observed standard deviations. In fact, the RMSEs for the standard deviations were lower than the RMSEs for the means and skews. This supports the conclusion that the relatively low correlations for the standard deviations reflect a restriction in range rather than an inability of the model to explain systematic variability. According to the RMSEs, the instantiation model does an excellent job of predicting the standard deviations. Furthermore, across studies, the correlations between predicted and observed standard deviations were quite respectable and significant in most cases. Thus, for all of these reasons, it is clear that the instantiation principle is implicated in the results for standard deviations, not just in the results for means and skew.

GENERAL DISCUSSION

The instantiation principle predicts that detailed information about a category's exemplars enters into the category's representation. To assess this prediction, a parameter-free exemplar model that incorporated detailed distributional information about categories was fit to the typicality judgments of human subjects. Of interest was whether the model's simulated distributions of typicality judgments would do a good job of predicting the observed distributions. Across three studies, high fits were observed for means (correlations generally above 0.9), and respectable fits were observed for standard deviations (generally between 0.6 and 0.9), with there being good reason to believe that extraneous factors, such as a restricted range of variation, limited the model's ability to predict standard deviations. In addition, Study 3 found a high fit for the skews of the distributions (0.87), as well as a deterioration in prediction as atypical exemplars were dropped from the simulations. Across studies, the same general patterns occurred for different superordinates (*animal*, *food*), for complex categories as well as simple ones at two taxonomic levels (e.g. *small reptile*, *small animal*), and for shifting typicality gradients across modified superordinates (*animal*, *small animal*, *dangerous animal*).

Together, all of these results indicate that detailed knowledge of exemplars entered into subjects' representations of categories. Rather than representing these categories only with features that are generally true across exemplars, subjects appeared to include detailed and idiosyncratic information from individual exemplars. The ability of the instantiation model to fit subtle distributional properties of typicality judgments is impressive, given that the model is parameter free and does not attempt to model the entire typicality task. Indeed, for natural categories, we know of no other model in the literature that comes close to accounting for as much of the variance in typicality judgments or any other categorisation task.

Further Support for the Instantiation Principle

Some additional support for the importance of instantiation can be found in previous research. For example, findings from Rosch et al. (1976) suggest that basic level instantiations of superordinate taxonomic categories should be easier to process than the superordinates themselves. For instance, Rosch et al. found that the members of superordinate categories do not share a common shape, whereas their instantiations at the basic level do. Similarly, superordinate categories convey relatively few common features, whereas basic categories convey substantially more. These results suggest that it might be easier to process the instantiations of superordinate categories than to process the general categories directly. Indeed, Rosch et al. reported slower verification times for superordinates than for basic level categories, which can be interpreted as reflecting the added step of instantiation for superordinates.

Studies on reading comprehension provide well-known evidence of instantiation. After reading the sentence ''The fruit was made into wine'', the word ''grape'' serves as a better retrieval clue for this sentence than the word ''fruit'' (Anderson et al., 1976). Similarly, Roth and Shoben (1983) found that subjects were faster to read the word ''cow'' when a previous sentence described milking an animal, compared to when a previous sentence described riding an animal. On reading ''animal'', subjects instantiated it as a cow or a horse depending on the context. Finally, McKoon and Ratcliff (1989) reported evidence for an instantiation process in recognition memory. For example, subjects were likely to falsely recognise the word ''cow'' if they had previously read ''animal'' in the context of milking. Garnham (1985) provides a further review of instantiation in memory and comprehension tasks.

The instantiation principle has also been applied successfully to reasoning. Osherson et al. (1991) suggested that people perform instantiation when evaluating the strength of inductive arguments such as:

$$\frac{\text{Canines have sesamoid bones.}}{\text{Mammals have sesamoid bones.}}$$

According to Osherson et al., people instantiate the categorical terms in such arguments (e.g. *canine*, *mammal*) to evaluate the generalisability of the predicted property (e.g. *sesamoid bones*). Osherson et al. tested this proposal by collecting subjects' instantiations of these categories and incorporating them into their model's predictions. In the present example, they found that the argument's strength could be predicted by the similarity between instantiations of *canine* (e.g. *dog*, *wolf*) and instantiations of *mammal* (e.g. *dog*, *human*). In addition, for a quite different domain of reasoning, Glass and Waterman (1988) argued that people's predictions about the entertainment value of movies reflected an instantiation process. Using a method analogous to that of Osherson et al. (1991), Glass and Waterman found that subjects evaluated brief descriptions of movies by retrieving previous instances of movies that fit the descriptions. This finding indicates that reasoning about uncertain future events depends critically on retrieving past instances. Finally, Tversky and Kahneman's

(1973) classic research on the availability heuristic strongly suggests that the retrieval of instances is central to many kinds of judgments.

Exemplar models of categorisation strongly embody the instantiation principle (Estes, 1994; Heit, 1992, 1994; Lamberts, 1994, 1995; Medin & Schaffer, 1978; Nosofsky, 1984). These models share the common assumption that various category judgments result from retrieving memory traces corresponding to members of the category. (In this way, exemplar models of categorisation are quite compatible with multiple-trace models of memory, e.g. Estes, 1994; Gillund & Shiffrin, 1984; Heit, 1993; Hintzman, 1988; Smith & Zaraté, 1992; see Jones & Heit, 1993, for a review.) Notably, most previous experiments supporting exemplar theory have relied on artificial categories learned by subjects in the laboratory, such as categories of geometric figures or fictional diseases. The present research is a step towards extending exemplar theory to natural categories such as *animal*, already known to the subject population (see also Heit, 1994). Research on natural categories and research on artificial categories has often proceeded on separate paths (see Murphy, 1993, for further discussion). The compatibility between the instantiation principle and exemplar models of categorisation suggests that judgments about natural categories and artificial categories can be explained within a unified framework.

Finally, the instantiation principle is consistent with proposals by the philosopher George Berkeley (1710/1986). Berkeley criticised previous philosophical work on abstract ideas because he could not conceive of an abstract concept without conceiving of a specific instance. For example, he reported that he could not think of *triangle* or of *person* without thinking of a particular triangle or a particular person. The instantiation principle provides a descriptive account of one way that people make judgments about abstract categories. We do not take a position on Berkeley's stronger claim that it is *impossible* to conceive of general categories without conceiving of more specific instances.

In summary, there is much evidence in the literature to implicate the instantiation principle in human cognition.

Implementing the Instantiation Principle

Thus far, we have remained agnostic on how human cognition might implement the instantiation principle, primarily because our data only bear on the principle itself and not on its implementation. Nevertheless, it is important to consider the potential ways in which the cognitive system could implement this principle. Clearly, one possibility is to implement the instantiation principle in exemplar models, much like the instantiation model used here to test the instantiation principle. Less obvious is the possibility of implementing the instantiation principle in abstraction models. In the next two sections, we explore issues associated with these two forms of implementation.

Instantiation in Abstraction Models. Standard prototype models track the frequencies of independent features for a category, only including features in the category's prototype that lie above some threshold of frequency. As a result, idiosyncratic features of exemplars, whose frequency is typically low, fall below the threshold and fail to enter the prototype. Furthermore, because only the frequencies of independent features are tracked, combinations of features never enter the prototype either. The net result is the failure of prototypes to include detailed information about exemplars or small subsets of exemplars.

However, a variety of abstraction models have been proposed over the years that are more sensitive to detailed information in exemplars. For example, the models of Reitman and Bower (1973), Neumann (1974), and Hayes-Roth and Hayes-Roth (1977) all assume that abstracted summaries represent categories, yet further assume that these summaries include idiosyncratic features from exemplars, as well as feature combinations. Similarly, many connectionist models, which are closer to abstractionist models than to exemplar models, also have considerable ability to abstract idiosyncratic features (e.g. McClelland & Rumelhart, 1985) and correlated features (e.g. through the hidden units in multi-layer nets). As Barsalou (1990) demonstrates, the mechanisms in these particular abstraction models allow them to be informationally equivalent to exemplar models, or at least to approach equivalence, depending on the specific model. The important implication for this paper is that these types of abstraction models are fully able to implement the instantiation principle.

To see how these abstraction models implement the instantiation principle, consider their assumptions about encoding and retrieval. At encoding, all independent features from each exemplar may be stored, not just features above some threshold of frequency. Although storing such detailed information is storage intensive, it leads to better categorisation, because a greater amount of potentially diagnostic information is used (Barsalou & Hale, 1993). Besides encoding all independent features in exemplars, these abstraction models also encode the feature combinations in each exemplar. For example, these models track the frequencies of all feature pairs across exemplars, all feature triples, and so forth. Although storing the entire power set of features across exemplars can lead to storage problems, there are reasonable assumptions that abstraction models can make a store feature combinations conservatively, yet still provide considerable computational power (Barsalou, 1990). Thus, these abstraction models implement the instantiation principle because their encoding mechanisms track the frequencies of idiosyncratic features and at least some feature combinations.

At retrieval, these abstraction models could implement the instantiation principle in a variety of ways. Analogous to how most exemplar models retrieve the entire exemplar set in making a category judgment, abstraction models could retrieve the entire set of features and feature combinations. Indeed, this is how the models of Reitman and Bower (1973), Neumann (1974), and Hayes-Roth and Hayes-Roth (1977) work. Alternatively, small subsets of features could be retrieved that vary across subjects and vary from situation to situation (Barsalou, 1987, 1989, 1990). One subject might retrieve one subset of features and feature combinations from the knowledge abstracted for a category, whereas another subject might retrieve a different subset. In this way, different subjects instantiate the same category in different ways, drawing on detailed knowledge from the entire exemplar distribution stored in memory.

Instantiation in Exemplar Models. In contrast to abstraction models, exemplar models instantiate category representations with memories of exemplars, not with features abstracted over them. As in the instantiation model explored here, different subjects' representations of the same category vary, because different exemplar memories instantiate these representations. The instantiation model clearly outlines how an exemplar version of the instantiation principle could be implemented.

Nevertheless, there are a number of issues that could be considered in developing the model further. One important issue concerns the assumption that people evaluate exactly one instantiation for each category. In research on artificial categories, exemplar models typically retrieve all known category exemplars rather than a single exemplar (although it is difficult to distinguish between these two possibilities empirically, as noted by Heit, 1992, and Medin & Schaffer, 1978, but see Barsalou, Lamberts, & Huttenlocher, in prep., and Nosofsky & Palmeri, 1995). Likewise, for natural categories, it is an open question whether people retrieve only one instantiation or some set of instantiations. Perhaps the number of instantiations varies widely, from one on some occasions, to several on others.

When performing typicality judgments, if subjects do retrieve multiple instantiations for a category, then their judgments should be more stable than if they retrieve only a single instantiation, just as the standard error of a mean decreases as sample size increases. Analogously, if the instantiation model were to base each judgment on multiple instantiations, then the standard deviations it produces would become smaller, because they would now be computed over means of judgments rather than over individual judgments. Building this into the model might compensate for the model's tendency to overpredict standard deviations (see, especially, Figs. 4, 9, and 10). Thus, an important issue for future research is to examine the number of instantiations that people use to represent categories and then build this structure into the model.

Instantiation at Multiple Taxonomic Levels. Both exemplar and abstraction models must address an additional issue, namely, at what taxonomic level does instantiation occur? Do people instantiate all categories with detailed exemplar information? Or do some categories fail to follow the instantiation principle? In the typicality judgment task studied here, do subjects instantiate the superordinates on typicality judgments, as well as the subordinates, with detailed category information? When judging the typicality of *mammal* with respect to *animal*, for example, do subjects instantiate *animal* as well as *mammal*? Furthermore, when subjects retrieve an instantiation for either a superordinate or a subordinate, do they instantiate the instantiation as well? If so, how far down does the instantiation process go? For example, if subjects instantiate *mammal* with *dog*, do they then instantiate *dog* with a specific breed and/or specific individuals?

For the most part, the subordinates in our studies were from the superordinate level of taxonomies, in the sense that the members of these categories do not overlap much in shape or characteristic features (Rosch et al., 1976). For example, the members of *reptile*, such as *alligator*, *snake*, and *lizard*, do not share a common shape or have the same characteristic features. Because we found substantial evidence of instantiation for these categories, we can conclude that at least Roschian superordinate categories exhibit instantiation. However, our materials also suggest that the instantiation principle applies to basic level categories as well. Two of the subordinates we used for *animal*, *bird* and *fish*, are widely believed to reside at the basic level (Rosch et al., 1976). For example, many instantiations of *bird*, such as *robin*, *parakeet*, and *pigeon*, are similar in shape and other characteristics. In analyses of the 18 distributions for *bird* and *fish* in Study 3, the instantiation model was successful at predicting the means, $r = 0.92$, $P < 0.001$, and the standard deviations, $r = 0.65$, $P < 0.01$. Thus, it appears that the instantiation process is not limited to superordinate taxonomic categories but extends at least to some basic level categories as well.

Should it turn out that people instantiate categories at multiple taxonomic levels, this raises the issue of how multiple sets of instantiations are compared. To see this, imagine that when people judge the typicality of *mammal* in *animal*, they instantiate *animal* with *horse*, *trout*, and *eagle*, and instantiate *mammal* with *dog*, *horse*, and *bear*. Alternatively, if people are using more abstract knowledge, they may instantiate *animal* with *alive*, *eats*, and *reproduces*, and instantiate *mammal* with *fur*, *bear live young*, and *mammary glands*. For either exemplar or abstraction models, how are these two sets of instantiations compared? Do subjects evaluate the average, minimum, or maximum pairwise similarity between these sets? Osherson et al. (1991) found that the maximum similarity between sets of instantiations played an important role in induction. In general, however, it remains to be seen how this comparison process proceeds across a wide variety of conceptual tasks.

In summary, the evidence from our studies, as well as related results from other researchers, favours the inclusion of the instantiation principle as one of the central principles describing conceptual processing. However, much remains to be learned about how far the instantiation principle extends across conceptual domains and tasks, about how this process is implemented in cognitive mechanisms, and about how this process relates to the other conceptual mechanisms that complement it.

Manuscript received 13 October 1995
Manuscript accepted 5 December 1995

REFERENCES

Anderson, J.R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Anderson, R.C., Pichert, J.W., Goetz, E.T., Schallert, D.L., Stevens, K.V., & Trollip, S.R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, *15*, 667–679.

Armstrong, S.L., Gleitman, L.R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 263–308.

Barsalou, L.W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp.101–140). Cambridge: Cambridge University Press.

Barsalou, L.W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp.76–121). Cambridge: Cambridge University Press.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T.K. Srull & R.S. Wyer (Eds.), *Advances in social cognition* (Vol.3, pp.61–88). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, & M.A. Conway (Eds.), *Theories of memory* (pp.29–101). London: Lawrence Erlbaum Associates Ltd.

Barsalou, L.W., & Hale, C.R. (1993). Components of conceptual representation: From feature lists to recursive frames. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analyses* (pp.97–144). San Diego, CA: Academic Press.

Barsalou, L.W., Lamberts, K., & Huttenlocher, J. (in prep.). *Processing individuals during categorization*.

Barsalou, L.W., & Sewell, D.R. (in prep.). *Flexibility and stability in category representations*.

Berkeley, G. (1710/1986). *A treatise concerning the principles of human knowledge.* La Salle, IL: Open Court.

Estes, W.K. (1994). *Classification and cognition.* Oxford: Oxford University Press.

Garnham A. (1985). *Psycholinguistics: Central topics.* New York: Routledge.

Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Glass, A.L., & Waterman, D. (1988). Predictions of movie entertainment value and the representativeness heuristic. *Applied Cognitive Psychology*, *2*, 173–179.

Hampton, J.A. (1993). Prototype models of concept representation. In I. Van Mechlen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp.67–88). San Diego, CA: Academic Press.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, *16*, 321–338.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*, 341–380.

Heit, E. (1993). Modelling the effects of expectations on recognition memory. *Psychological Science*, *4*, 244–252.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1264–1282.

Heit, E., & Barsalou, L.W. (1992). An instantiation model of category typicality and instability. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, (pp.360–365). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Jones, C.M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 799–812.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1003–1021.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*, 161–180.

Marr, D. (1982). *Vision.* San Francisco: W.H. Freeman.

McClelland, J.L., & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

McKoon, G., & Ratcliff, R. (1989). Inferences about contexually defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1134–1146.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Murphy, G.L. (1993). A rational theory of concepts. In G.V. Nakamura, R. Taraban, & D.L. Medin (Eds.), *The psychology of learning and motivation: Categorization by humans and machines* (pp.329–359). San Diego, CA: Academic Press.

Neumann, P.G. (1974). An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, *2*, 241–248.

Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Nosofsky, R.M. (1992). Exemplars, prototypes, and similarity rules. In A.M. Healy, S.F. Kosslyn, & R.M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes*, (pp.149–167). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Nosofsky, R.M., & Palmeri, T.J. (1995). *An exemplar-based random walk model of speeded classification* (Technical Report No.142). Cognitive Science Program, Indiana University.

Osherson, D.N., Stern, J., Wilkie, O., Stob, M., & Smith, E.E. (1991). Default probability. *Cognitive Science*, *15*, 251–269.

Reitman, J.S., & Bower, G.H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, *4*, 194–206.

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.

Roth, E.M., & Shoben, E.J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, *15*, 346–378.

Smith, E.E., & Medin, D.L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E.R., & Zaraté, M.A. (1992). Exemplar-based models of social judgment. *Psychological Review*, *99*, 3–21.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

APPENDIX A
PRODUCTION FREQUENCIES FOR STUDY 2

Beverage

Coke	12
beer	6
milk	4
water	3
soda pop	6
Diet Coke	2
Diet Pepsi	
Dr. Pepper	
ginger ale	
grapefruit juice	
Kool-aid	
Jack Daniels	
Pepsi	

Dairy Product

milk	32
cheese	4
ice cream	4

Dessert

ice cream	16
cake	7
cheesecake	4
pie	3
chocolate cake	2
chocolate mousse	2
apple pie	
chocolate	
chocolate pudding	
eggs	
mousse	
strawberry shortcake	

Fish

salmon	14
trout	8
cod	4
catfish	2
scrod	2
bass	
bluegill	
halibut	
herring	
fluke	
lox	
perch	
swordfish	
Whaler sandwich	
whitefish	

Fruit

apple	23
orange	5
banana	3
peach	3
pear	3
mango	
plum	
watermelon	

Meat

steak	18
beef	10
hamburger	3
ham	2
bacon	
chicken	
ground beef	
pork chop	
red meat	
sausage	
sirloin	

Poultry

chicken	35
turkey	2
bacon	
Cornish hen	
eggs	

Seasoning

pepper	8
oregano	7
salt	7
garlic	6

basil	2
dill weed	2
A1 steak sauce	
cinnamon	
cloves	
curry	
paprika	
salad dressing	
seasoned salt	
soy sauce	

Vegetable

carrot	13
broccoli	9
green bean	4
celery	3
spinach	2
tomato	2
banana	
bean sprout	
cauliflower	
eggplant	
lettuce	
potato	
string bean	

APPENDIX B
PRODUCTION FREQUENCIES FOR STUDY 3

Amphibian

frog	13
lizard	8
salamander	4
turtle	4
toad	3
snake	2
alligator	
crocodile	
duck	
fish	
salmon	
spider	

Bird

robin	9
eagle	6
parakeet	3
pigeon	3
blue jay	2
canary	2
cardinal	2
parrot	2
sparrow	2
bluebird	
condor	
crow	
mockingbird	
pelican	
penguin	
pterodactyl	
swallow	
yellow-bellied sapsucker	

Fish

trout	6
piranha	4
goldfish	3
guppy	3
perch	3
salmon	3
carp	2
minnow	2
shark	2
tuna	2
barracuda	
blowfish	
bluefish	
catfish	
eel	
frog	
mahi-mahi	
pike	
sardine	
sunfish	

Insect

bee	6
fly	6
spider	5
ant	4
cockroach	4
mosquito	4
beetle	3
butterfly	2
grasshopper	2
caterpillar	
Japanese beetle	
ladybug	
tiger	

Mammal

whale	9
human	8
dog	6
bear	2
cat	2
elephant	2
monkey	2
platypus	2
dinosaur	
dolphin	
giraffe	
horse	
kangaroo	
rabbit	

Micro-organism

amoeba	15
bacteria	4
cell	3
paramecium	3
algae	2
virus	2
E. coli	
euglena	
fungus	
gastropod	
gnat	
hookworm	
larva	
maggot	
placenta	
protozoa	
yeast	

Reptile

alligator	9
snake	9
lizard	8
crocodile	3
iguana	3
newt	2
turtle	2
brontosaurus	
frog	

Small Amphibian

frog	14
salamander	6
lizard	5
newt	3
iguana	2
tadpole	2
baby alligator	
baby crocodile	
chameleon	
crow	
fish	
goldfish	
snail	
turtle	

Small Bird

hummingbird	12
sparrow	9
canary	3
robin	3
chickadee	2
chicken	2
finch	2
cardinal	
cockatiel	
meadowlark	
nuthatch	
parakeet	
parrot	
wren	

Small Fish

goldfish	16
guppy	9
minnow	5

perch	3
bluegill	
chub	
neon fish	
piranha	
sunfish	
trout	
tuna	

Small Insect

ant	19
fly	7
mosquito	4
flea	3
beetle	2
grasshopper	2
cockroach	
ladybug	
mite	

Small Mammal

mouse	9
rat	4
dog	3
human	3
cat	2
platypus	2
rabbit	2
shrew	2
baby whale	
bird	
bunny rabbit	
chipmunk	
duck	
gopher	
kitten	
koala bear	
mole	
monkey	
opossum	
squirrel	

Small Micro-organism

amoeba	17
bacteria	8
cell	3
algae	2
E. coli	2
flea	2
cryptosporida	
paramecium	
plankton	
protozoa	
spirochete	
zygote	

Small Reptile

lizard	15
snake	6
chameleon	5
iguana	3
turtle	3
frog	2
alligator	
anole	
baby alligator	
baby lizard	
garden snake	
salamander	

Dangerous Amphibian

crocodile	6
poison toad	6
alligator	3
Gila monster	3
shark	3
snake	3
frog	2
lizard	2
newt	2
rattlesnake	2
giant iguana	
horned toad	
kimono dragon	
salamander	
snapping turtle	
tick	
toad	
turtle	

Dangerous Bird

eagle	9
hawk	9
vulture	6
pterodactyl	3
condor	2

crow	2
falcon	2
ostrich	2
bald eagle	
finch	
seagull	
sparrow	

Dangerous Fish

piranha	21
shark	14
barracuda	2
electric eel	
lionfish	
puffer	

Dangerous Insect

mosquito	8
bee	7
spider	6
killer bee	3
cockroach	2
tarantula	2
wasp	2
black widow	
bumblebee	
centipede	
honeybee	
locust	
praying mantis	
scorpion	
tick	
tsetse fly	
yellow jacket	

Dangerous Mammal

tiger	8
lion	7
bear	6
human	6
whale	3
shark	2
buffalo	
gorilla	
great white shark	
kodiak bear	
leopard	
orangutan	
panther	
weasel	

Dangerous Micro-organism

bacteria	11
virus	9
amoeba	4
AIDS virus	3
cold virus	
flea	
flu virus	
germ	
giardia	
hookworm	
leech	
mycoplasma	
paramecium	
spermatozoa	
tapeworm	
tick	

Dangerous Reptile

alligator	14
snake	9
crocodile	3
boa constrictor	2
cobra	2
iguana	2
dinosaur	
Gila monster	
lizard	
python	
rhinoceros	
tiger snake	
tyrannosaurus	