Relational Versus Absolute Representation in Categorization

DARREN J. EDWARDS and EMMANUEL M. POTHOS
Swansea University

AMOTZ PERLMAN
Bar-Ilan University, Israel

This study explores relational-like and absolute-like representations in categorization. Although there is much evidence that categorization processes can involve information about both the particular physical properties of studied instances and abstract (relational) properties, there has been little work on the factors that lead to one kind of representation as opposed to the other. We tested 370 participants in 6 experiments, in which participants had to classify new items into predefined artificial categories. In 4 experiments, we observed a predominantly relational-like mode of classification, and in 2 experiments we observed a shift toward an absolute-like mode of classification. These results suggest 3 factors that promote a relational-like mode of classification: fewer items per group, more training groups, and the presence of a time delay. Overall, we propose that less information about the distributional properties of a category or weaker memory traces for the category exemplars (induced, e.g., by having smaller categories or a time delay) can encourage relational-like categorization.

The problem of how naive observers represent information is a fundamental one in psychology. On one hand, there is a strong intuition that psychological representations have to be veridical descriptions of the physical and perceptual properties of the stimuli in our environment. The bulk of the modeling work in categorization involves such representations. For example, both exemplar theory (Medin & Schaffer, 1978; Nosofsky, 1984, 1986) and prototype theory (Homa & Vosburgh, 1976; Posner & Keele, 1968; Reed, 1972) typically formulate predictions in terms of items represented in a way that directly corresponds to their actual physical properties. However, it seems that the representational capacity of human cognition is a lot more flexible than that.

On the other hand, there has been an abundance of evidence that representation can be guided by the perceptual information for a stimulus, relative to other stimuli. For example, an influential tradition of relevant evidence comes from research on analogical reasoning. Analogical mapping is a process of comparison to identify shared relations between two knowledge systems, such as two objects. The generated comparisons are thought to play a role in relational reasoning (Gentner, 1983, 1989; Gick & Holyoak, 1980, 1983; Holyoak & Thagard, 1995), in learning and using rules (Anderson & Lebiere, 1998; Lovett & Anderson, 2005), in the appreciation of perceptual similarities (Medin, Goldstone, & Gentner, 1993), and in the production of language, science,

American Journal of Psychology
Winter 2012, Vol. 125, No. 4 pp. 481–497 • © 2012 by the Board of Trustees of the University of Illinois

mathematics, and art. Research on analogical mapping shows how, when one is comparing objects such as an elephant, truck, mouse, and ball, abstract shared properties can be identified that capture key relations between the objects (e.g., elephant and truck are both big; mouse and ball are both small).

We can call relational-like (we will justify the use of this grammatically awkward construction shortly) a categorization process in which items are represented in terms of some relational, abstract property (e.g., "small vs. large"). A relational classification is therefore based on a relational property, which is independent of the particular physical properties of individual exemplars and rather depends on the relations between exemplars. The implied converse mode of categorization, absolute-like categorization, involves item representations that veridically correspond to the actual physical properties of the items (e.g., "approximately 6 cm vs. approximately 20 cm").

A categorization researcher can ask whether there might be circumstances that spontaneously lead to a preference for a more absolute-like or relational-like mode of categorization. In this respect, prior research is uninformative. Most studies either assume one form of representation or demonstrate that a particular form of representation is plausible (e.g., in analogical reasoning, the objective is commonly to demonstrate situations in which analogies can be used to solve reasoning problems). However, it is rarely the case that alternative possible representations for the same stimuli have been directly contrasted within the same paradigm.

This discussion immediately leads to an important methodological problem: How can a researcher determine whether a particular categorization reflects absolute-like or relational-like representations? In a typical manipulation in this work, participants see items varying along a single dimension in a training phase. The training items are organized into two categories, a category of small items and a category of large items. In test, suppose that there are four test items, two of which are smaller than the others but also such that they are all larger than the items in the "large" training category (Figure 1). It seems straightforward to assume that a relational-like categorization would mean that the Category A exemplars are represented as smaller than the Category B ones, so that the two shorter of the four test stimuli will be classified as Category A instances and the two larger as Category B ones. By contrast, with an absolute-like categorization, all four test instances should be classified as Category B instances because their physical properties are more similar to those of the Category B members.

Two important qualifications underwrite the robustness of this paradigm. First, an assumed relational-like representation is *not* the same as a fuzzy absolute-like representation. If the representation of the test instances is absolute-like but inexact in some sense, then they should still be classified in Category B, as long as the difference between Category A and Category B exemplars is large enough (see Figure 1). This can be arranged in a straightforward way, for example in an experiment where we have exemplars in training Category A approximately 30 mm long and exemplars in training Category B 60 mm long, whereas test exemplar sizes are 80 mm and 120 mm long. In this case, an absolute-like representation would yield a classification in which both test items would be classified into Category B. Second, one can assume that the default response bias of participants would be to select some test instances as members of one training category and other instances as members of the other. Such a response bias clearly favors a relational-like mode of categorization in our experiments. Crucially, the conclusions we are seeking to derive in this work are not whether a particular manipulation leads to absolute-like or relational-like categorizations, but rather whether it leads to *more* absolute-like or relational-like categorization, in relation to a baseline manipulation.

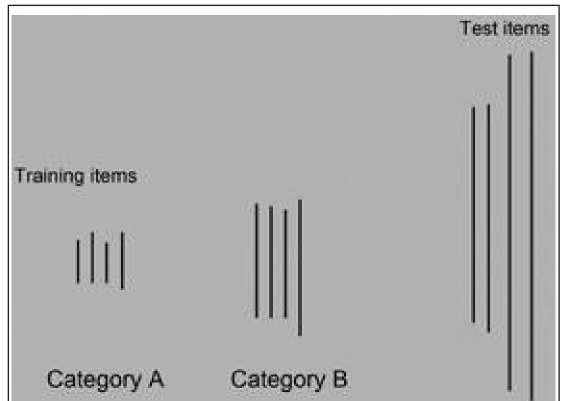

FIGURE 1. Typical manipulation in the present work, in which each line corresponds to a stimulus. Stimuli vary along a single dimension, overall length

These considerations also justify our use of the terms *relational-like* and *absolute-like* instead of just *relational* or *absolute:* In any single experiment, a pattern of responses can look more or less relational compared with the pattern of classification against some baseline condition. However, it is not possible to decide whether the responses in any particular experiment correspond to a relational or absolute representation.

This discussion has some parallels with the research tradition on absolute identification tasks. In such tasks, a participant is presented with several stimuli of varying magnitudes along a particular dimension of physical variation, such as height. They are then asked to remember these stimuli and to place them in order from smallest to largest from memory. Using this task, Miller (1956) observed that people found it difficult to identify a particular item from a set of items that vary along a single dimensional continuum (e.g., length, brightness of color, or pitch of tone). Stewart, Brown, and Chater (2005; see also Brown, Marley, Donkin, & Heathcote, 2008, Laming, 1984, 1997) suggested that in absolute identification tasks a response is generated through a comparison between the current stimulus and the previous stimulus. In other words, response generation is assumed to be driven by "difference" information and not by information about the absolute physical properties of the stimuli. Relatedly, in sequential classification paradigms, there has been evidence that trial-to-trial information is used in a systematic way. For example, the similarity of consecutive stimuli may bias their classification (Jones, Love, & Maddox, 2006), or the perception of sequentially presented stimuli may be affected by how they are classified (Zotov, Jones, & Mewhort, 2010; cf. Ward & Lockhead, 1971).

Such results also reveal situations in which the perception of an individual item is a function of the context in which it is presented. However, they concern the systematicity with which physical differences between the current stimulus and the previous stimulus affect the perception of the current stimulus and therefore differ from the relational-like categorization (or representation) we outlined earlier. In general, there is little doubt that presentation order can affect perception and categorization, whether this concerns individual stimuli (Jones et al., 2006; Zotov et al., 2010) or more general practice effects and related changes in the overall representational bias (e.g., Johansen & Palmeri, 2002). However, in the present study we are interested in whether the distributional properties of training and test items in a categorization task can bias toward relational-like or absolute-like classification. Such biases appear to be independent of biases relating to presentation order. Therefore, the focus of our empirical design has been on concurrent presentation, so as to avoid complications from sequential effects and enable the pursuit of a more specific research objective.

What determined whether an observed categorization is more absolute-like or relational-like? In categorization, it seems implausible that there are not circumstances in which the representations we use are absolute (e.g., see Goldstone, 1994; Goldstone & Barsalou, 1998). So the question becomes, Under what circumstances might we expect that an absolute classification mode will be preferred? It is reasonable to suggest (e.g., in terms of a minimalist Bayesian intuition) that the absolute properties of a category (e.g., information about particular exemplars, as would be required by exemplar theory, or information about a prototype, as would be required by prototype theory) would be inferred with more confidence if more category exemplars were studied in the training phase. In other words, suppose that the correct hypothesis about the absolute value of the prototypes of the two presented categories is such that the first prototype is P1 and the second prototype is P2.

Let us label the training exemplars of the categories as D1, D2, D3, etc. We suggest that $P(P1, P2|D1, D2, D3)$ would be lower than $P(P1, P2|D1, D2, D3, D4, D5, D6 \ldots)$. In other words, it would be possible to evaluate with more confidence a hypothesis about the absolute properties of the category prototypes if more training exemplars are processed. Likewise, if there were four categories there would be four values to infer about the physical values of the prototypes, P1, P2, P3, P4. A straightforward extension of this reasoning suggests that $P(P1, P2|D1, D2, D3)$ would in general be higher than $P(P1, P2, P3, P4|D1, D2, D3)$. In other words, we would need more information to support or reject a more complex hypothesis (about the physical properties of category prototypes). This argument does not assume that participants represent categories with exemplars or prototypes. Rather, our claim is that a representation of a category based on absolute physical values of the training exemplars is more likely to be possible (and hence, we predict, preferred by the cognitive system) if there are more training exemplars per category.

This approach leads to some straightforward predictions: More training exemplars per category should lead to more absolute-like classification. More categories should lead to more confusion between the particular physical properties of each category and therefore to less absolute-like classification. Finally, other manipulations, which undermine participants' confidence in the absolute physical properties of the training items, would also lead to less absolute-like classification (we used one such manipulation, a time delay). Note that these statements immediately raise a question: What exactly is the process for the transition between an initially relational representation and an absolute representation, as more exemplars are presented to participants? Intuitively, it seems that there is a threshold of evidence above which an initially relational representation becomes absolute. However, equally, the transition from relational to absolute representation may be more graded. Our design is not suited to answering this question. Rather, our design simply assumes that there is a probability for more relational versus more absolute representation, which can vary across our manipulations. Regardless of the exact psychological process, we can then explore, across participants, whether particular manipulations increase or decrease the probability for relational and absolute representations.

These predictions depend on a particular, incidental mode of category learning (cf. Pothos & Chater, 2002; Pothos et al., 2011). In our tasks, the training items were shown to participants in bundles corresponding to their intended categorizations (all items were presented concurrently). Thus, participants could readily perceive the training items in terms of their intended categories. In addition, participants were exposed to the training exemplars of each category briefly; they did not have an incentive (and were not encouraged) to memorize the exemplars or study them thoroughly. Indeed, because they had the training exemplars available in test there would be no reason for them to do so.

Such a mode of category learning contrasts with the more common supervised categorization method. In such cases, participants are exposed over a large number of trials to the same training exemplars repeatedly, until they can (typically) perfectly reproduce their classifications. Moreover, the category structures typically learned in this way are complex: Extensive training is needed before learning can be achieved (by contrast, in our experiments participants were shown very simple category structures). What would be the relevant expectations under such circumstances? Clearly, the fewer the exemplars, the easier participants would find learning the required classifications, and therefore the more salient would participants find the taught exemplar–classification label associations. Overall, the fewer the exemplars in a supervised categorization paradigm, the more pronounced we would expect exemplar effects to be. Indeed, this is exactly what has been found in previous work (Rouder & Ratcliff, 2004; Shin & Nosofsky, 1992).

Therefore, with the incidental category learning paradigm approach we predict that fewer exemplars typically lead to less absolute-like classification; however, with a standard supervised categorization paradigm fewer exemplars lead to more absolute-like classification (more pronounced exemplar effects). To reiterate, the key difference between the two approaches is that in the latter case, participants have no choice but to represent the training exemplars in an absolute way. They receive typically dozens of training trials (e.g., in Rouder & Ratcliff's Experiment 2, there were 96 trials in a block and 10 blocks in a session, with each participant going through two sessions of training) in which they have to learn to associate particular exemplars with particular labels. In our case, by contrast, the training exemplars never really have to be learned (they are present throughout the experiment, and the category structure is very simple). Additionally, the test exemplars could be classified by interpreting the training exemplars in two radically different ways.

To further demonstrate the conceptual difference between relational and absolute representation, we used the generalized context model (GCM; Nosofsky, 1984, 1986). The GCM is a model widely used for the study of categorization processes. According to the GCM, the categorization of a new exemplar is determined by the similarity between that new exemplar and those stored in memory. The GCM sums the similarity of a new item with the items in each category and predicts that the new item will be classified in the category for which this summed similarity is greatest. For example, a new instance will be classified as belonging to Category A rather than Category B if it is more similar to the A exemplars than the B exemplars. More specifically, exemplars are represented in a multidimensional space; each exemplar is stored together with its category label. In a simple, one-dimensional case, the distance (d_{ij}) between two stimuli S_i and S_j is given as

$$d_{ij} = |x_i - x_j| \tag{1}$$

where x_i is the absolute magnitude of S_i and x_j is the absolute magnitude of S_j. For an *m*-dimensional space, the weighted Minkowski power formula is used, so that the distance (d_{ij}) between stimuli S_i and S_j is given as

$$d_{ij} = \left[\sum_m w_m \cdot |x_{im} - x_{jm}|^r\right]^{1/r} \tag{2}$$

In Equation 2, x_{im} denotes the value of exemplar i on psychological dimension m. The r value defines the distance metric of the psychological space. For example, the city block metric is defined with $r = 1$, and the Euclidean distance metric is defined with $r = 2$. Shown in Equation 2 are also the attention weight parameters w_m, which model the degree to which a participant attends to a particular dimension. The similarity between stimuli S_i and S_j is a function of their distance. Similarity is typically a monotonically decreasing function of distance, as in

$$\eta_{ij} = e^{-cd_{ij}^q}, \tag{3}$$

In Equation 3, η_{ij} is the similarity between S_i and S_j, where $q = 1$ leads to an exponential function and $q = 2$ leads to a Gaussian function. The sensitivity parameter, c, determines how quickly the similarity between stimuli S_i and S_j is reduced with distance.

The probability of classifying stimulus S_i in Category A is proportional to the similarity between S_i and all the A exemplars, as in Equation 4; in that equation, the β_A parameters are category biases, which indicate whether there might be a prior bias to identify new items as being members of a particular category.

$$H_{iA} = \beta_A \sum_{xj \in C_A} \eta_{ij}, \tag{4}$$

Finally, the actual probability of making a Category A response given stimulus S_i, when there are two alternative categories (A and B), is given by Equation 5.

$$P(R_A|S_i) = \frac{H_{iA}}{H_{iA} + H_{iB}}, \tag{5}$$

At present, the GCM can be used to explore whether our proposed design can capture a distinction between relational-like and absolute-like representation, as we intend. Consider a set of training items that vary in terms of overall height, such that there is a category of Chomps (heights: 32, 35, 36, and 40 mm) and a category of Blibs (heights: 62, 64, 66, and 70 mm). Next, consider test items with heights of 81, 85, 121, and 124 mm. Clearly, the Chomps are smaller than the Blibs. Also, two of the test items are smaller than the other two, but the test items are all more similar to the Blibs than the Chomps (cf. Figure 1).

Suppose that a participant classified all the test items into the Blibs category. We expect this to indicate that the representation of the items was absolute-like. Does the GCM confirm this expectation? We first examined the GCM on the basis of the absolute representation of the stimuli, to find a sum of squares value of 0 (smaller sum of squares values indicate better fit). We then introduced a relational representation, which encoded for the basic fact that there were two small and two large stimuli in the training set (encoded as 1 and 5, respectively) and two small and two large stimuli in test also (encoded in the same way). This encoding clearly captures the spirit of the relational representation, in that two items are small or large, relative to the other immediately relevant items, and irrespective of the absolute physical properties. As expected, on the basis of this representation, the GCM was unable to predict a classification of all the test stimuli into the Blibs category (sum of squares value: 2)

Conversely, suppose that a particular participant classified the two smaller test items in the Chomps category and the two larger ones in the Blibs category. Such a classification pattern is plausibly associated with relational-like mode of representation. When examining GCM predictions on the basis of the absolute representation of the stimuli, the GCM was unable to predict the required classification (sum of squares value: 4). By contrast, using a relational representation, the GCM could perfectly predict the required performance pattern (sum of squares value: 0). Note that the results with the relational representation do not depend on the particular coding scheme, and, for example, the same conclusions are obtained (sum of squares values of 2 or 4) if the small and large items are encoded as, say, 5 and 10 respectively, or 1 and 2.

The GCM is thus able to confirm that one particular pattern of classification is most consistent with an absolute classification and another one with a relational classification. Of course, the model cannot a priori predict whether it is more likely that participants will adopt an absolute or relational classification. To study the circumstances when this is more likely is the purpose of the present study. To briefly summarize our hypotheses, we suggest that when it is difficult to derive accurate training category representations (prototypes or exemplars) based on the physical properties of the studied objects, the cognitive system is more likely to use relational-like representations. Thus, we expect more relational-like categorization when there are fewer items per group, when there are more groups, or when there is a time delay between the initial presentation of training stimuli and test items. For the last hypothesis, we suggest that the time delay will deteriorate the memory of the specific exemplar representation and therefore reduce the available information about the distributional properties, leading to a greater likelihood of relational-like representation.

EXPERIMENT 1

Experiment 1 provides a baseline examination of the basic experimental design. There were two training categories, a category of Chomps and the category of Blibs, and four test items (the Appendix shows the heights for the stimuli in all experiments). If participants adopted an absolute-like mode of categorization, then all four test items should be classified in the category of Blibs (that is, all the large items would be classified in the same category). By contrast, if participants adopted a relational-like mode of classification, the two smaller test items should be classified with the training category of smaller items (Chomps), and the two larger test items should be classified with the category of larger training items (Blibs). Note that with a single experiment, it is impossible to gain insight into the circumstances under which absolute-like or relational-like representations are more likely to occur. For example, results of this experiment were driven partly by a propensity to represent the stimuli in a relational-like or absolute-like way but also by possible task demands, such as a bias to assign some of the test stimuli to all the available training categories. Experiment 1 is the baseline manipulation. We were interested in whether the additional manipulations in subsequent experiments lead to a shift in favor of relational-like or absolute-like representations, relative to Experiment 1. By comparing the results of subsequent experiments with those of Experiment 1, we effectively factor out such possible task demands.

METHOD

Participants

A total of 63 Swansea University students took part in the experiment for a small payment. Participants were tested individually and were all experimentally naive.

Materials

Twelve items were created using Corel Draw (Figure 2). Each item was presented on a card and consisted of a picture of a flower grounded on a solid base. The picture of the flower consisted of a yellow bud with eight petals and a blue stem. Eight items, grouped into two categories, made up the training stimuli. We selected stimuli that had a naturalistic character, so as to encourage participants to engage with the task.

The group of Chomps consisted of four flowers that were of the following heights: 32, 35, 36, and 40 mm. The group of Blibs consisted of four flowers

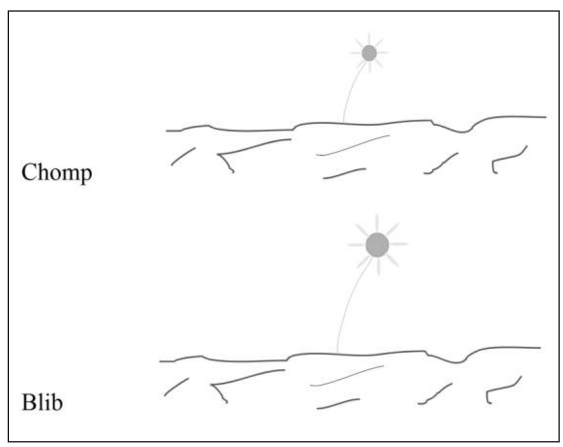

FIGURE 2. Sample stimuli. The top image is an example of an item belonging to the Chomps category (in Experiment 1), and the bottom image is an example of the Blibs category

with greater heights: 62, 64, 66, and 70 mm. There were four test items of flowers with heights of 81, 85, 121, and 124 mm (see the Appendix). The items in the category of Chomps and the items in the category of Blibs were such that height differences between any successive items were computed assuming a Weber fraction of 8%. The three authors independently verified that each stimulus could readily be discriminated from all others. Note that because participants saw all stimuli concurrently, even small differences between stimuli could be detected. Note also that the width of the flowers would increase in size, as would the overall height. Regarding the empirical findings, it does not matter whether participants attended to just the height of the stimuli, just the size of the flowers, or the configural dimension of overall stimulus size; the conclusion about possible absolute-like or relational-like categorization would be the same. Accordingly, we report height only values, because these provide the easiest way to label the stimuli.

Procedure

Participants were first presented with written instructions to the effect that they were about to see some items that belonged to two imaginary categories (called Chomps and Blibs) and that the experimenter would tell participants which items went in which categories. At that point, participants were shown the two groups of stimuli; each item was presented on a single card. All the training items, from both groups, were always presented simultaneously in our experiments, so that the items from each group were arranged in distinct piles (the items in each group were counterbalanced). Participants were asked to look through every item in the two groups in their own time (this was typically less than 3 min, and they typically went through the items one at a time but were not told specifically to do so). The training items then remained with the participant for the duration of the experiment.

Subsequently, participants were presented with new written instructions, indicating that new items would be shown and that each participant had to decide for each new item whether it was a Chomp or a Blib. The instructions stated, "There are no right or wrong answers! You have to classify each item as a Chomp or a Blib." After the presentation of the instructions, participants were presented with the four test items simultaneously. Participants typically needed a few minutes to make their responses, which they did by placing each of the test items into the training categories that they wanted to categorize these with.

RESULTS AND DISCUSSION

We define absolute-like categorization to correspond to a response pattern whereby all test items were considered Blibs, and in relational-like categorization the two smaller test items were considered Chomps and the two larger test items Blibs. Directly analogous definitions for absolute-like and relational-like categorization were used in the other experiments as well. Absolute- (or relational-) like categorizations will be influenced partly by whether the training stimuli are represented in an absolute or relational way. Equally, the relative proportion of absolute-like or relational-like categorization will depend on other factors, such as whether there is a response bias to classify some test stimuli into all available training categories and the particular physical characteristics of the stimuli (this latter issue is considered in Experiment 2b). Therefore, we cannot say from the results of a single experiment whether absolute-like categorization constitutes evidence for absolute-like representation. This becomes possible by comparing the results of two or more experiments so that we can examine whether a particular manipulation increases the tendency for absolute-like categorization. Notwithstanding these issues, the characterization of participant responses as absolute-like and relational-like seems like a good starting point in considering our data.

The responses of some participants were such that they did not conform to this characterization. Such participants were eliminated from the analyses here and elsewhere, because their results do not bear on the hypotheses we are interested in. Of course, if our experimental design works as intended, we would expect that few participants would produce such in-between responses. In this experiment, only one participant was eliminated because he or she categorized the stimuli in a way that did not clearly fit our definitions of absolute-like and relational-like representation. Forty-eight participants adopted a relational-like categorization mode and 14 an absolute-like one. Note that the motivation for requiring participants to classify each test item as a Chomp or a Blib was precisely so that classification patterns would inform as much as possible the contrast between relational-like and absolute-like classifications that we are interested in.

Here and elsewhere we adopted chi-square twotailed tests to examine any preference for relational- versus absolute-like classification of the test items, against what would be expected by chance or in relation to results from other experiments. A chi-square test against chance simply examines whether the proportion of absolute-like categorizations is equal to that of relational-like categorizations. We assume that a straightforward 50–50 observed-chance split is most appropriate for the chi-square analysis, as we were interested in testing against the null hypothesis that an absolute-like classification is equally likely to a relational-like classification. However, as noted, it is not the observed versus chance analysis that is particularly interesting but rather the comparison of classification performance across experiments.

In Experiment 1 there was a highly significant tendency for participants to prefer a relational classification, against chance, $\chi^2(1) = 10.08$, $p = .001$. Overall, Experiment 1 demonstrates the baseline condition and the analytical approach. It is clear that several participants (14 out of 62) did not feel obliged to assign the test instances to all the training categories. However, the preference for relational-like classifications in Experiment 1 was probably partly because of task demands. Accordingly, in itself, the conclusion from Experiment 1 is not interesting. In subsequent experiments, by altering the key characteristics of Experiment 1 and observing participants' performance, we will present a series of results that support our hypothesis about when relational-like or absolute-like categorization is more likely to occur.

EXPERIMENT 2A

In Experiment 1 we showed that relational-like categorization can be observed when each of the training categories had four test items per training group (Chomps and Blibs). In Experiment 2a we doubled the number of items per training group. We suggest that this may encourage absolute-like classification. It is possible that when more evidence is available about the distributional properties of the training items, concrete (absolute) information about the category exemplars (or prototypes) would be more available and so it would be such information that drives the classification of new exemplars. Thus, increasing the number of exemplars per category should increase absolute-like categorization.

METHOD

Participants, Materials, and Procedure

Fifty-nine Swansea University students took part in the experiment for a small payment. Participants were tested individually and were all experimentally naive (here and elsewhere, no participant took part in more than one of the present experiments). Materials consisted of the same two groups of flower images (Blibs and Chomps) but with eight instead of four items in each group. The heights of the members of the Chomps category were 35, 36, 40, 42, 44, 46, 47, and 49 mm, and the heights of the members for the Blibs category were 62, 64, 66, 70, 74, 75, 76, and 77 mm (see the Appendix). The test items and procedure were the same as in Experiment 1.

RESULTS AND DISCUSSION

In this experiment, we observed 25 participants providing relational-like classification of the test items and 32 absolute-like classifications (2 participants were eliminated because their responses could not be characterized as absolute-like or relational-like). Examining participants' pattern of responding against chance, as before, did not identify a preference for relational-like or absolute-like classification, $\chi^2(1) = 0.43$, $p = .51$. Crucially, when we compare the results of Experiment 2a with the results of Experiment 1, there was a highly significant interaction, $\chi^2(1) = 14.10$, $p = .0002$. This result indicates that in Experiment 2a participants were more likely to adopt an absolute-like mode of categorization than those in Experiment 1. Such a conclusion supports the hypothesis outlined in motivating Experiment 2a.

In sum, increasing the number of exemplars per category increased the preference for representing the stimuli in an absolute-like way, as one would predict by assuming that more concrete information about the training items increases absolute-like categorization.

EXPERIMENT 2B

We suggested that the finding of more absolute representation in Experiment 2a than in Experiment 1 can be attributed to the fact that there were more items per group in Experiment 2a. However, there is an alternative, potentially contributing factor. First of all, let us define *difference* as the difference in magnitude between the largest item in a small training category and the smallest item in the next large training category. Note that there is evidence in the literature that increasing the physical differences between stimuli increases discriminability, which, presumably, encourages absolute representation (Alluisi & Sidorsky, 1958; Braida & Durlach, 1972; Eriksen & Hake, 1955; Pollack, 1952).

The difference value for Experiment 1 is calculated as (Largest Chomp – Smallest Blib)/(Average size of largest Chomp – Smallest Blib) = (62 – 40)/51 = 0.43. Likewise, the difference value for Experiment 2a is 0.23. Hence, in these experiments we found that a *decrease* in difference was associated with more absolute representation. One can speculate that this is the case because having better-separated training categories encourages a more categorical representation of these categories.

Is it difference or the number of items per group that is the primary cause for more absolute representation in Experiment 2a? We designed this experiment to be as closely matched as possible to Experiment 2a but with the difference value matched to Experiment 1.

METHOD

Participants, Materials, and Procedure

Fifty-four Swansea University students took part in the experiment for a small payment. Participants were tested individually and were all experimentally naive (here and elsewhere, no participant took part in more than one of the present experiments). Materials consisted of the same two groups of flower images (Blibs and Chomps) but with eight instead of four items in each group. The heights of the members of the Chomps category were 32, 34, 35, 35, 36, 37, 38, and 40 mm, and the heights of the members for the Blibs category were 62, 63, 64, 65, 66, 68, 69, and 70 mm. Accordingly, the difference value in this experiment is (62 – 40)/51 = 0.43, which is identical to that of Experiment 1. The test items and procedure remained the same as in Experiment 1.

RESULTS AND DISCUSSION

In this experiment, we observed 42 participants providing relational-like classification of the test items and 8 participants producing absolute-like classifications (4 participants were eliminated because their responses could not be characterized as absolute-like or relational-like). Examining participants' pattern of responding against chance, we identified a preference for relational-like classification, $\chi^2(1) = 13.07$, $p = .0003$. When we compared the results of Experiment 2b with the results of Experiment 1, there was no significant interaction: $\chi^2(1) = 0.76$, $p = .38$.

The results of Experiment 2b indicate that doubling the items per group is in itself insufficient for inducing a preference for absolute judgment. Apparently, the physical differences between the stimuli in the different training groups (as quantified by our difference measure) plays an important role as well. We return to this issue when considering the comparison between Experiments 3 and 4.

EXPERIMENT 3

In Experiment 3 we provided a manipulation that was intended to weaken the concreteness of information for the training items and so (if our reasoning is correct) promote relational-like categorization. Accordingly, in this experiment we doubled the number of training groups, from two to four (the four training categories were called Chomps, Blibs, Zlogs, and Glabs). The Chomps had heights of 30, 29, 36, and 30 mm (an item was repeated); the Blibs 38, 39, 40, and 41 mm; the Zlogs 47, 48, 49, 50 mm; and the Glabs 58, 61, 62, and 63 mm. It can therefore be seen that the heights of the members of three of the four categories conformed to the simple ordering smallest, small, large, and largest. The heights of the four test items were 48, 62, 113, and 183 mm. Accordingly, one test item was the same size as one Zlog, and another test item was the same size as one Glab; the other two test items were larger than all the training items.

The fact that two test items were identical to two training items might plausibly increase absolute-like classification. However, if our suggestion that the concreteness of information for the training exemplars increases absolute-like classification is correct, then the converse prediction is made: With more categories, one would expect that participants would be more confused about the exact physical attributes of the members of each category, so that relational-like classification would be favored.

Note that Lacouture, Li, and Marley (1998) demonstrated that increasing the number of possible responses reduces the frequency of correct responses in absolute identification tasks. The manipulation in this experiment relies on a similar assumption, namely that with more categories, increased confusability between stimulus–category label associations would lead to more relational-like categorization. However, in Lacouture et al.'s work there was a single normative response. That is, all responses were either correct or wrong, and participants had to represent the stimuli in an absolute way for correct responses to be possible. By contrast, in our experiments participants were not constrained to represent the stimuli in a specific way. They could represent them in a presumably more absolute or more relational way. Our manipulation in Experiment 3 exactly follows the logic of Lacouture et al. By providing more response categories, we assumed that participants would find it more difficult to adopt absolute representations. For example, this difficulty might relate to higher confusability between stimulus–category label associations. Note that we were not interested in the difficulty with which multiple categories can be learned and, so, during the test phase, participants could observe all the training stimuli correctly arranged into their respective categories.

METHOD

Participants, Materials, and Procedure

A total of 79 Swansea University students took part in the experiment for a small payment. Materials consisted of the same flower images but with four instead of two groups (labeled Chomps, Blibs, Zlogs, and Glabs) and with four items for each group. The heights of all the stimuli are given in the Appendix. The procedure for this experiment was the same as for Experiment 1, except that participants were trained with four categories and were required to categorize each of the four test items into any of the four groups (whereas in Experiment 1 there were only two groups). The instructions stated that any possible classification of each test item was possible. Also, in this experiment participants were told more clearly that any possible classification was possible. Specifically, instructions before the test phase were augmented with the following text: "Any possible classification is allowed. The test items could be classified between all four possible categories, between three of the possible categories, between two of the possible categories or just one of the categories."

RESULTS AND DISCUSSION

In this case, relational-like categorization corresponded to grouping the smallest test item with the training group with the smallest members (Chomps), the second smallest test item with the training group with the second smallest members (Blibs), etc. An absolute-like classification corresponded to assigning test item 48 in the category of Zlogs (because there was a Zlog of the same size) assigning test item 62 to the category of Glabs (because there was a Glab of the same size) and, finally, assigning the two remaining test items (113 and 183) to the category with the largest members (the category of Glabs). Fifty-nine participants adopted a relational-like categorization mode and 9 an absolute-like categorization mode.

Responses from 11 participants were removed from the data because they did not fit the definitions of relational-like or absolute-like categorization and so do not bear on the hypothesis tested. Note that the greater number of rejected participants in this experiment is a natural implication of the combinatorics of category assignment in test. In Experiments 1, 2, and 2b, there were two test items and two possible category assignments, for a total of four possibilities. By contrast, in Experiment 3 there were four test items and four category assignments, for a total of 256 possibilities. The greater number of possibilities inevitably led to greater noise in category assignments. Note, finally, that the responses from the rejected participants did not appear to conform to a particular consistent pattern. The same points apply to Experiment 4.

We used a chi-square test to investigate whether relational-like or absolute-like categorization was preferred (against chance) in classifying the test items into the categories of Chomps, Blibs, Zlogs, and Glabs. There was a tendency to form relational-like categorizations in this experiment, $\chi^2(1) = 21.25$, $p < .0005$. We next examined whether in Experiment 3 there were more relational-like categorizations than in Experiment 1. There was an interaction in this direction, but it did not prove to be significant, $\chi^2(1) = 1.95$, $p = .163$. Note that because we were exploring specific hypotheses (e.g., that increasing the number of exemplars per group would bias toward absolute-like classification), a Bonferroni correction is not appropriate (see Nakagawa, 2004, or Perneger, 1998, for general criticisms of the Bonferroni correction and related adjustments to the significance threshold).

EXPERIMENT 4

In Experiment 3 we doubled the number of training groups and so observed a preference for relational-like classification, although this preference was not higher than that in Experiment 1. This preference should be reduced by doubling the number of members of each group if our hypothesis about relational-like versus absolute-like classification is correct.

METHOD

Participants, Materials, and Procedure

A total of 60 Swansea University students took part in the experiment. Materials consisted of the same flower images with four groups (labeled Chomps, Blibs, Zlogs, and Glabs) but with eight instead of four items for each group. These were the same training groups as in Experiment 3 (Chomps, Blibs, Zlogs, and Glabs), but each category in this case had eight members instead of four. The eight Chomps were flowers with heights of 27, 28, 28, 31, 30, 29, 24, and 21 mm; the eight Blibs had heights of 38, 39, 40, 41, 39, 40, 41, and 36 mm; the Zlogs had heights of 47, 48, 49, 50, 49, 50, 51, and 52 mm; and the Glabs had heights of 58, 61, 62, 63, 61, 60, 62, and 63 mm. The four test items were the same as in Experiment 3 and had heights of 48, 62, 113, and 183 mm. The procedure for this experiment was the same as for Experiment 3. Note that some exemplars were repeated so as to fit the required number of stimuli within the available range of physical variation.

RESULTS AND DISCUSSION

In this experiment, we observed 36 participants as categorizing according to relational-like categorization and 14 according to absolute-like categorization. The responses of 10 participants were removed because they did not fit the definitions of absolute-like or relational-like categorization. As was the case in Experiment 3, the rejected responses did not appear to conform to a consistent pattern of classification. We first examined whether there was any evidence for a preference of absolute-like or relational-like categorization against chance. A chi-square test showed that participants preferred a relational-like mode of classification, $\chi^2(1) = 5.07$, $p = .024$. The crucial comparison regarding our hypothesis corresponds to whether there were a greater proportion of absolute-like categorizations in Experiment 4, compared with Experiment 3. This was indeed the case: There were significantly more absolute-like categorizations in Experiment 4 than in Experiment 3, as predicted, $\chi^2(1) = 4.0$, $p = .045$.

So, in increasing the number of stimuli per group from four (Experiment 3) to eight (Experiment 4), we observed a shift for more absolute-like classification. This situation is analogous to that concerning the comparison between Experiments 1 and 2 except that we showed (in Experiment 2b) the greater difference value in Experiment 1 to partly account for the preference for relational-like classification. Regarding the comparison between Experiments 3 and 4, there is no such complication. We computed a difference value for Experiment 3 as the average of the difference values between all successive (in terms of average physical size) category pairs, which came out to be $0.11 = (0.05 + 0.14 + 0.15)/3$. The analogous value for Experiment 4 was $0.13 = (0.15 + 0.14 + 0.11)/3$. Thus, in Experiment 4 we observed the shift for absolute-like classification, relative to Experiment 3, even though the difference measures for these two experiments were matched.

We also compared Experiment 4 with Experiment 2a to examine whether increasing the number of groups would have led to an increase in the number of relational categorizations (note that the number of items in each training group was the same in Experiment 4 and Experiment 2a). For this comparison, we found $\chi^2(1) = 8.61$, $p = .003$, which supports our hypothesis that increasing the number of groups increases the preference for relational-like categorizations (see also Lacouture et al., 1998).

Experiments 1–4 examined our hypothesis in terms of manipulating the number of training categories and the size of each category. In all cases, our results were consistent with a general hypothesis about preference for absolute-like or relational-like classification, according to which, when it is possible to derive more concrete information about a category, then absolute-like classification should be favored. In Experiment 5 we attempt an alternative test of this hypothesis.

EXPERIMENT 5

Experiment 5 is based on Experiment 2a (two training categories, with eight items per group). The difference between the two experiments is that instead of asking participants to classify the test items immediately after presentation of the training items, we asked them to return 1 week later and then make their classification decisions. According to our hypothesis, the time delay should deteriorate the memory traces for the training items and thus increase the proportion of relational-like categorizations. In other words, if participants are unable to remember the exact physical characteristics of the stimuli, they might be more likely to attempt to classify the test stimuli on the basis of relational features, such as "small versus large."

Note that our intuition here regarding the impact of time on representation appears reasonable in the context of the limited training participants had with the stimuli (effectively, a single exposure). However, plausibly, the impact of time on representation may be different after extensive training. Likewise, it is worth noting that there is evidence that in absolute identification tasks participants revert to more absolute representations, after a delay (e.g., Ward, 1987; Ward & Lockhead, 1971; see also Weber, Green, & Luce, 1977). But, as noted, there are important differences between the notion of relational-like and absolute-like representation that we use in this work and that of relative representation, as they are relevant in absolute identification tasks.

METHOD

Participants, Materials, and Procedure

A total of 59 Swansea University students took part in the experiment for a small payment. The materials were identical to those in Experiment 2a. Briefly, Chomps had heights of 35, 36, 40, 42, 44, 46, 47, 49 mm, Blibs had heights of 62, 64, 66, 70, 74, 75, 76, 77 mm, and the test items had heights of 81, 85, 121, and 124. The procedure was likewise identical to that of Experiment 2a, except that participants were asked to make their classifications a week after they had studied the training items. Moreover, in Experiment 5, classification of the test items took place without having the training items available.

RESULTS AND DISCUSSION

In this experiment, we observed 47 participants classifying the test items in a relational-like way and 5 classifying the test items in an absolute-like way. Results from 7 participants were removed because their classifications could not be characterized as absolute-like or relational-like. In Experiment 5 there was a strong tendency (against chance) to form relational-like categorizations, $\chi^2(1) = 20.27, p < .0005$. We next compared the results of Experiment 5 with the results of Experiment 2a to find that in the former the proportion of relational-like categorizations was much higher, as predicted: $\chi^2(1) = 26.25, p < .0005$. This is consistent with a hypothesis such that a time delay causes decay in memory and therefore weakens the memory for the absolute-like (physical) properties of the training exemplars, which encourages a relational-like classification.

GENERAL DISCUSSION

In categorization, it is sometimes the case that representation appears to involve abstract features. For example, the sun is "large," a cheetah is "fast," and a Christmas dinner is "plentiful." In such and similar cases, it is uncommon to provide a more specific (absolute) impression of the corresponding characteristics. These are exactly the kind of situations we are trying to model with our experiments, that is, situations when absolute characterizations appear inappropriate or inconvenient, in some sense, and so people resort to more relational representations. Equally, there is a strong intuition that categorization and representation are often absolute-like, that is, more tightly linked to the physical properties of the relevant stimuli.

What are the circumstances when a relational-like representation mode might be preferred to an absolute one? The extensive research tradition on analogical reasoning has made extensive use of relational features to understand analogical reasoning. However, this work does not provide any prediction of whether absolute-like or relational-like representations are more likely to be adopted in a categorization task. With six experiments, we aimed to provide some boundary conditions on this issue. Our results show that the cognitive system appears to adopt a fairly principled and adaptive way of preferring relational-like to absolute-like representations, at least with respect to classifying novel instances into incidentally taught categories.

Experiment 1 was the baseline manipulation, in which more relational-like categorizations were observed than absolute-like ones. The predominance of relational-like categorizations could result from a greater bias for relational-like categorization as such, but it might also result from participants wanting to assign some test instances to all the available categories (i.e., a task demand). Conclusions about absolute-like or relational-like categorization are possible only if the results of at least two experiments are compared.

Experiments 2a and 2b differed from Experiment 1 in that the former had twice as many stimuli per group as the latter. However, a shift for absolute-like classification was observed only in Experiment 2a, which had a *lower* difference value than Experiment 1. Thus, with only two categories, it appears that both increasing the number of items per group and decreasing the difference value can lead to absolute classification. The importance of increasing the number of items per group for absolute classification was also demonstrated by the comparison between Experiments 3 and 4, as the former was designed with four items per category and the latter eight; in both cases there were four training categories. Even though difference values were matched between Experiments 3 and 4, we observed the expected shift for absolute-like classification in Experiment 4. Experiment 5 was designed to have the same materials as Experiment 2a, but in that experiment we instead asked participants to classify the test items 1 week after the training items were presented. This manipulation led to a shift toward relational-like categorization when we compared Experiment 5 with Experiment 2a.

The results can be summarized in the following way (Table 1). First, when the number of items per category was increased, we sometimes observed a shift toward absolute-like categorization. Presumably, more exemplars per category imply that there is more information on the basis of which the cognitive system can represent a category in a concrete (absolute) way. However, we need to qualify this conclusion: Increasing the number of items per category led to a shift for absolute classification when there were four training categories (the comparison between Experiments 3 and 4), but when we had only two training categories, a shift for absolute classification was observed only when the difference value was also decreased (Experiments 1, 2a, and 2b).

Second, increasing the number of category groups enhances a relational-like mode of categorization. In this case, we suggest that the cognitive system finds it more difficult to concurrently keep track of the distributional properties of several categories (and so represents them in a relational-like way). Accordingly, when there is a requirement to learn more categories, it adopts a simpler, relational-like way of representing category information. Finally, when we introduced a time delay, relational-like classification was encouraged. It appears that, at least in some cases, a delay leads to a less detailed mode of representing the information from the training phase, so that the cognitive system would abandon an absolute-like mode of representation and instead prefer a representation in terms of less specific relational features, such as "small versus large."

There is a consistent theme underlying all our findings. When it comes to categories composed of simple, schematic stimuli, the cognitive system appears to operate like a standard statistical engine: The more the information about the distributional properties for the exemplars of each category, the more likely it is that the category will be represented in a way that directly relates to the physical properties of the stimuli (cf. Ashby & Maddox, 1993; Chater, 1999; Lee & Vanpaemel, 2008; Tenenbaum & Griffiths, 2001). Distributional category information could be undermined in a variety of ways: by having fewer items per category, more categories (which would lead to more confusion about which item belongs to which group), a time delay, and the particular physical differences between the stimuli in the different training categories. Our experiments involved manipulations of all these factors apart from the last one and support the impression that all these factors are relevant in establishing expectations for relational-like versus absolute-like categorization.

Our findings most directly affect categorization theory. In brief, there are two classes of models: models of supervised categorization (e.g., prototype or ex-

TABLE 1. Summary of Results

Experiment	Classification frequency	Predominant classification	Items per group	Number of training groups	Time delay	Significance (all χ^2 tests with 1 df)
Experiment 1 vs. chance	48 R 14 A	Relational	4	2	None	.001
Experiment 2a vs. chance	25 R 32 A	Nonsignificant	8	2	None	.51
Experiment 2a vs. 1	48 R 14 A vs. 25 R 32 A	Absolute	4 vs. 8	2 vs. 2	None	.0002
Experiment 2b vs. chance	42 R 8 A	Relational	8	2	None	.0003
Experiment 2b vs. 1	42 R 8 A vs. 48 R 14 A	Nonsignificant	4 vs. 8	2 vs. 2	None	.38
Experiment 3 vs. chance	59 R 9 A	Relational	4	4	None	<.0005
Experiment 3 vs. 1	59 R 9 A vs. 48 R 14 A	Nonsignificant	4 vs. 4	4 vs. 2	None	.163
Experiment 4 vs. chance	36 R 14 A	Relational	8	4	None	.024
Experiment 4 vs. 2a	36 R 14 A vs. 25 R 32 A	Relational	8 vs. 8	4 vs. 2	None	.003
Experiment 4 vs. 3	59 R 9 A vs. 36 R 14 A	Absolute	4 vs. 8	4 vs. 4	None	.045
Experiment 5 vs. chance	47 R 5 A	Relational	8	2	1 week	<.0005
Experiment 5 vs. 2a	47 R 5 A vs. 25 R 32 A	Relational	8 vs. 8	2 vs. 2	1 week vs. none	<.0005

emplar theory; Nosofsky, 1984; Homa & Vosburgh, 1976) and models of unsupervised categorization (e.g., the rational model of Anderson & Betz, 2001, or the simplicity model of Pothos & Chater, 2002). Supervised categorization models typically operate on a default representation of the stimuli but have the ability to transform this representation, typically through attentional parameters (attentional parameters effectively select out a subspace of the default representational space). Unsupervised categorization models can sometimes predict the dimensions participants will spontaneously prefer when categorizing a set of stimuli (Pothos & Bailey, 2009; Pothos & Close, 2008).

Could our results be explained within such modeling frameworks? The principles that guide dimensional selection in supervised categorization models have to do with identifying the representation, which makes the required categorization easiest to learn. One could conceivably propose that the representation of the stimuli in our experiments is made of both absolute and relational properties. Then, the categorization task in test makes one set of properties more useful than the other. There are two problems with this approach. First, in our experiments it appears that the emphasis on relational properties (e.g., in Experiment 1 vs. 2a) has to do with the processing of both the training items and the test items. By contrast, supervised categorization models set their parameters only during the processing of the training stimuli. Second, there is an infinite number of possible relational properties. How could we, as modelers, decide a priori which are the appropriate relational properties to use in an experimental situation? Similar considerations apply in the case of models of unsupervised categorization (noting, in any case, that dimensional selection for such models is less well developed than that for supervised categorization models).

In sum, our results show situations in which classification would indicate relational-like or absolute-like representation, in a way that takes into account distributional information from both the training and the test stimulus sets. Such flexibility in category representation is difficult to reconcile with current categorization models and represents an exciting avenue for further research.

APPENDIX. LENGTH (MM) OF EACH ITEM IN EACH GROUP

Experiment 1			Experiment 2a			Experiment 2b			Experiment 3					Experiment 4					Experiment 5		
Chomp group	Blib group	Test group	Chomp group	Blib group	Test group	Chomp group	Blib group	Test group	Chomp group	Blib group	Zlog group	Glab group	Test group	Chomp group	Blib group	Zlog group	Glab group	Test group	Chomp group	Blib group	Test group
32	62	81	35	62	81	32	62	81	30	38	47	58	48	27	38	47	58	48	35	62	81
35	64	85	36	64	85	34	63	85	29	39	48	61	62	28	39	48	61	62	36	64	85
36	66	121	40	66	121	35	64	121	36	40	49	62	113	28	40	49	62	113	40	66	121
40	70	124	42	70	124	35	65	124	30	41	50	63	183	31	41	50	63	183	42	70	124
			44	74		36	66							30	39	49	61		44	74	
			46	75		37	68							29	40	50	60		46	75	
			47	76		38	69							24	41	51	62		47	76	
			49	77		40	70							21	36	52	63		49	77	

NOTE

Address correspondence about this article to Darren J. Edwards, Department of Psychology, Swansea University, Swansea, SA2 8PP, England (e-mail: D.J.Edwards@swansea.ac.uk).

REFERENCES

Alluisi, E. A., & Sidorsky, R. C. (1958). The empirical validity of equal discriminability scaling. *Journal of Experimental Psychology, 55,* 86–95.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review, 8,* 629–647.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483–502.

Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. J. (2008). An integrated, principled account of absolute identification. *Psychological Review, 115,* 396–425.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology, 52A,* 273–302.

Eriksen, C. W., & Hake, H. W. (1955). Absolute judgments as a function of stimulus range and number of stimulus and response categories. *Journal of Experimental Psychology, 49,* 323–332.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). London, England: Cambridge University Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65,* 231–262.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Johansen, M. K., & Palmeri, T. J. (2002). Representational shifts in category learning. *Cognitive Psychology, 45,* 482–553.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 316–332.

Lacouture, Y., Li, S., & Marley, A. A. J. (1998). The roles of stimulus and response set size in the identification and categorization of unidimensional stimuli. *Australian Journal of Psychology, 50,* 165–174.

Laming, D. R. J. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical and Statistical Psychology, 37,* 152–183.

Laming, D. R. J. (1997). *The measurement of sensation.* London, England: Oxford University Press.

Lee, M. D., & Vanpaemel, W. (2008). Exemplars, prototypes, similarities and rules in category representation: An example of hierarchical Bayesian analysis. *Cognitive Science, 32,* 1403–1424.

Lovett, M. C., & Anderson, J. R. (2005). Thinking as a production system. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 401–429). New York, NY: Cambridge University Press.

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

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

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review, 63,* 81–97.

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology, 15,* 1044–1045.

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. (1986). Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments? *British Medical Journal, 316,* 1236–1238.

Pollack, I. (1952). The information of elementary auditory displays. I. *Journal of the Acoustical Society of America, 24,* 745–749.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Pothos, E. M., & Bailey, T. M. (2009). Predicting category intuitiveness with the rational model, the simplicity model, and the generalized context model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1062–1080.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303–343.

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107,* 581–602.

Pothos, E. M., Perlman, A., Bailey, T. M., Kurtz, K., Edwards, D. J., Hines, P., & McDonnell, J. V. (2011). Measuring category intuitiveness in unconstrained categorization tasks. *Cognition, 121,* 83–100.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General, 133,* 63–82.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of "dot-pattern" classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*(4), 881–911.

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

Ward, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 216–227.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73–78.

Weber, D. L., Green, D. M., & Luce, R. D. (1977). Effect of practice and distribution of auditory signals on absolute identification. *Perception & Psychophysics, 22,* 223–231.

Zotov, V., Jones, M. N., & Mewhort, D. J. K. (2010). Contrast and assimilation in classification and exemplar production. *Attention, Perception, & Psychophysics, 73,* 621–639.