Psychonomic Bulletin & Review
2007, 14 (5), 823-833

Similarity, distance, and categorization: A discussion of Smith's (2006) warning about "colliding parameters"

Daniel J. Navarro
University of Adelaide, Adelaide, South Australia, Australia

The idea that categorization decisions rely on subjective impressions of similarities between stimuli has been prevalent in much of the literature over the past 30 years and has led to the development of a large number of models that apply some kind of decision rule to similarity measures. A recent article by Smith (2006) has argued that these *similarity–choice* models of categorization have a substantial design flaw, in which the similarity and the choice components effectively cancel one another out. As a consequence of this cancellation, it is claimed, the relationship between distance and category membership probabilities is linear in these models. In this article, I discuss these claims and show mathematically that in those cases in which it is sensible to discuss the relationship between category distance and category membership at all, the function relating the two is approximately logistic. Empirical data are used to show that a logistic function can be observed in appropriate contexts.

If every stimulus in our world were perceived as an entirely unique object, people would be inundated with an immense amount of pointless information. So we organize objects into categories, allowing us to describe the world in a simpler manner and to generalize better to novel situations. Not surprisingly, then, understanding the nature of human concepts and the way in which they shape our categorization behavior has remained one of the central topics in cognitive psychology. Ever since the decline of the classical view of concepts, which assumed that a category could be held together via a collection of features both necessary and sufficient to determine category membership, one of the key ideas in psychological theories has been that a category can be held together by a loose family resemblance among objects. According to this *resemblance view* of categorization (Rosch, 1978), it is the similarities between things that govern the extent to which people judge an item to belong to a category.

When formalized as cognitive models (e.g., Medin & Schaffer, 1978; Nosofsky, 1984), similarity-based theories rely on two key assumptions. First, they assume that the subjective sense of similarity between items decreases very rapidly (exponentially, in fact) as the items are made more distant in some suitable sense. Second, in order to account for behavior in forced choice tasks, the models incorporate some kind of choice rule. In a recent article, Smith (2006) has argued that these *similarity–choice models* for categorization suffer from a major design flaw, resulting from a complex interaction between these two assumptions. In effect, he argued that similarity and choice "cancel," leaving a simple linear function relating category distance to category membership probabilities. The implication of the claim is that we might benefit by discarding the framework provided by similarity-based models of categorization and replacing them with models that predict category membership by applying linear functions to distances. Citing work by Roberts and Pashler (2000), Smith emphasized the importance of thinking about more than a model's data fit and the value of examining the internal structure of cognitive models, since it is just such an analysis that uncovers the cancellation effect.

In terms of the general suggestion that modelers need to be aware of complex interactions between parameters that can arise in some cases, it is difficult to disagree with Smith (2006). Indeed, in recent years, an extensive literature has built up regarding how to measure model performance in an appropriate way (e.g., Balasubramanian, 1997; Myung & Pitt, 1997; Navarro, Pitt, & Myung, 2004; Pitt, Myung, & Zhang, 2002) and how to understand the characteristic patterns that a model can produce (e.g., Myung, Kim, & Pitt, 2000; Pitt, Kim, Navarro, & Myung, 2006). Moreover, these methods have frequently been applied to the understanding of categorization models, including RULEX (Navarro, 2005), the generalized context model (Navarro, 2007), and ALCOVE (Pitt et al., 2006). At a more specific level, it is straightforward to demonstrate that the claim that similarity and choice "cancel" is incorrect. When the analysis is performed correctly, it becomes clear that, under the assumption that similarity does not dissociate from an appropriate measure of category distance, an approximate functional relationship does exist,

D. J. Navarro, daniel.navarro@adelaide.edu.au

but it is described by a logistic equation, not a linear one. When similarity and category distance dissociate, there is no guarantee that *any* function of category distance can adequately describe the behavior of these models.

The plan of this article is as follows. The first section provides an overview of similarity–choice models, which is followed by a simple example that demonstrates that a linear relationship does not hold in general. After this, I derive the logistic function that relates distance and category membership when the categories are "well behaved." This is followed by a demonstration showing that there is *no* simple relationship (i.e., neither linear nor logistic) between the two when the category structures become more complicated. After doing so, I discuss an empirical example in which the categories are well behaved and the nonlinear, logistic relationship is observed.

Similarity–Choice Models

In order to construct any mathematical model for how people make judgments about stimuli, we need a language that allows us to describe stimuli in a sufficiently precise manner. One common approach (e.g., Tversky, 1977) is to compile a list of characteristics possessed by an object. For instance, in order to provide a rich enough description of the coffee cup on my desk, I might note that it (1) is cylindrical, (2) is open at the top, (3) has a handle, (4) is blue, and (5) has "NIPS 2002" written on it. By way of comparison, the disposable cup sitting behind it is cylindrical and open at the top, but it is not blue, lacks a handle and does not have writing on the side. Since each of these characteristics is binary (the item either has it or does not), we refer to them as *binary features*. We would describe the coffee cup using the feature vector 11111 but would describe the disposable cup using the feature vector 11000. Any such description could be called a *stimulus representation*. Often, the "features" of a stimulus (e.g., its height) take on continuous values. A continuous-valued feature is often referred to as a *dimension*, and there are some subtleties that arise when switching between features and dimensions (discussed briefly in the Appendix). In a categorization context, we might consider a domain consisting of n stimuli, each of which can be described in terms of m features. The ith stimulus, then, is represented by the vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$, where x_{ik} denotes the value of stimulus i on feature k.

At its simplest, the idea behind resemblance views of categorization is that stimuli that people perceive as being highly similar will be much more likely to belong to the same category. Accordingly, for the ith and jth stimuli in the domain, a mathematical model requires that we define the *stimulus similarity*, s_{ij}, between the two. The difficulty, of course, is that it is not easy to define what we mean by "similar." As philosophers have pointed out (e.g., Goodman, 1972), it is very easy to end up in a situation in which the whole concept of similarity becomes circular. It is for this reason that psychological approaches to similarity are described in terms of a specific stimulus representation, and these representations are constrained by empirical data, generally via the use of multidimensional scaling, additive clustering, or other related methods (see Shepard, 1980).

One widely used approach for relating a stimulus representation to a subjective similarity relies on Shepard's (1987) law of generalization, which suggests that if there is some good way of measuring the *psychological distance*, d_{ij}, between the two items, similarity decays exponentially with distance:

$$s_{ij} = e^{-\lambda d_{ij}}. \tag{1}$$

In this expression, λ is a parameter that describes how quickly similarity decays with distance, sometimes referred to as the *sensitivity* parameter. Obviously, for this relationship to be useful, we need a good way of describing psychological distance. Although there are some complexities associated with how interstimulus distance should be measured (see the Appendix), for the present purposes it will suffice to add up the differences on each dimension:

$$d_{ij} = \sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|. \tag{2}$$

In order to make a judgment about category memberships, similarity–choice models construct a measure of the overall similarity between the to-be-classified item (TBCI) and the category as a whole. In a prototype model, a single idealized category member (the prototype) is constructed. Having done so, one calculates the distance from the prototype to the TBCI and, from that, the similarity. In an exemplar model, the similarity s_{iA} between stimulus i and category A is calculated by finding the similarity of stimulus i to each of the members of category A and simply taking the sum. To provide a sense of what these category similarity functions look like for an exemplar model, Figure 1 shows two simple categories, each consisting of three evenly spaced exemplars, differing only on the one

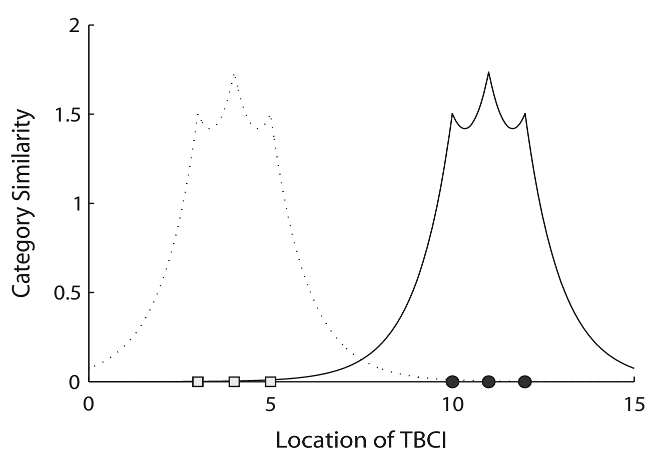

Figure 1. Category similarity functions for two simple one-dimensional categories, shown by the squares (category *A*) and circles (category *B*). The solid line shows the similarity to category *B*, and the dotted line shows the similarity to category *A*. For this figure, similarity to category is calculated using an exemplar model with $\lambda = 1$.

dimension. The plots show the similarity to each category for every possible location of the TBCI.

Irrespective of how category similarities are defined, the *choice rule* applied in these circumstances is a straightforward normalization (Luce, 1959).[1] If the learner is asked to choose between category A and category B, the choice rule implies that the probability $P(i \in A)$ with which item i is judged to belong to category A, rather than to category B, is simply

$$P\left(i \in A\right) = \frac{s_{iA}}{s_{iA} + s_{iB}}. \tag{3}$$

The membership probability $P(i \in A)$ is sometimes referred to as the *category endorsement level*.

Relating Distance to Category Membership

The discussion in Smith's (2006) article revolves around a "cancellation" phenomenon. The basic idea is that if we take a reasonable measure of the distance between the TBCI and the two categories and plot the difference between the two distances against the endorsement probability, we obtain a linear function. The article presents a number of simulations in which an approximately linear relationship is observed and concludes that "the cancellation of similarity by choice is mathematically inexorable and will always apply . . . even when distance and similarity dissociate" (Smith 2006, p. 748). If true, the cancellation hypothesis suggests that we might very well be able to replace the entire framework of resemblance and choice with a much simpler class of linear models. A number of findings might need to be reinterpreted in light of the new linear categorization models that would emerge. In view of the potential importance of this result, it is worth examining this "linear cancellation hypothesis" in some detail.

To provide a sense of the problem with the linear cancellation hypothesis, it is instructive to consider the two categories shown in Figure 1. The categories are very simple, since the stimuli vary only along a single dimension. Category A consists of three exemplars on the left, and category B consists of three exemplars on the right. If the cancellation phenomenon holds as generally as Smith's (2006) article implies, it would not be unreasonable to expect it to emerge if we apply an exemplar model to these categories. Unfortunately, it does not, as is illustrated by Figure 2. Applying much the same approach as Smith, we can trace out the functional relationship between average distance and endorsement that emerges for these two categories. In this figure, panel A shows the endorsement values for all possible locations of the TBCI, calculated by taking the category similarity functions in Figure 1 and applying the choice rule (Equation 3). In panel B, the average distance between the TBCI and the category members is calculated for all possible locations of the TBCI. Panel C then takes the difference in category distances (i.e., how much closer the TBCI is to category B than to category A) and plots it with the dashed line. To complete the demonstration, panel D plots the two functions against one another, giving us the relationship between distance and endorsement. Clearly, panel D does not depict a linear function and suggests a curvilinear function, such as a logistic or a cumulative normal ogive. Although a linear function with slope as a free parameter fits well to the 1,500 data points used to construct the function, accounting for 94.5% of the variance (i.e., $r^2 = .945$), a logistic function "fits" noticeably better, accounting for 99.99998% of the variance.[2] The superior performance in this case arises despite the fact that, as we will see, no free parameters were involved in "fitting" the logistic function. Although it would be unwise to draw strong conclusions from a single case, the existence of such a simple counterexample casts some doubt on the claim that the distance endorsement function is "inexorably" linear.

Deriving the logistic function. In order to understand why the relationship depicted in Figure 2 is accounted for so effectively by a logistic function, it is helpful to recognize that the variables plotted in panel D are $\delta_{iA} - \delta_{iB}$ on the x-axis and $P(i \in B)$ on the y-axis, where I have assumed that the category distance δ_{iA} is given by the average distance from the TBCI to the members of category A:

$$\delta_{iA} = \frac{1}{n_A} \sum_{j \in A} d_{ij} \;. \tag{4}$$

In this expression, n_A counts the number of items in category A. It is very important to recognize that this average distance measure is not explicitly used in similarity–choice models; it is for this reason that the symbol δ is used instead of d. Since δ is not part of the model, there is no guarantee that there is *any* particular r elationship between this category distance measure and the endorsement probabilities. However, the measure is fairly interpretable, and since the distance measures discussed by Smith (2006) tended to be the average distance δ_{iA} (Simulation 2), the total distance $n_A\delta_{iA}$ (Simulations 3, 5, 7, and 9), or a special case of one of these measures (Simulation 1), I will adopt it here.[3] To avoid unnecessary complication, it suffices for the moment to note that the following derivation relies on the *assumption* that $s_{iA} \approx e^{-\lambda\delta_{iA}}$ (a more general treatment is given in the Appendix). To the extent that this approximation holds (i.e., when category similarity and average distance do not dissociate), it is straightforward to substitute $e^{-\lambda\delta_{iA}}$ into Equation 3 and find the logistic function that relates average distance to category endorsement:

$$\begin{aligned} P\left(i \in A\right) &= \frac{s_{iA}}{s_{iA} + s_{iB}} \\ &\approx \frac{e^{-\lambda\delta_{iA}}}{e^{-\lambda\delta_{iA}} + e^{-\lambda\delta_{iB}}} \\ &= \frac{1}{1 + e^{-\lambda\left(\delta_{iB} - \delta_{iA}\right)}}. \end{aligned} \tag{5}$$

In the initial example, this approximation explained 99.99998% of the variance with λ fixed at the true value of 1; for all intents and purposes, the approximation is exact in this case. The nonlinearity of Equation 5 makes clear that it is simply incorrect to make any general claim that similarity and choice cancel to produce a linear relationship between distance and endorsement.

In view of the simplicity of this derivation, it should come as no surprise that it is not a new result; an almost identi-

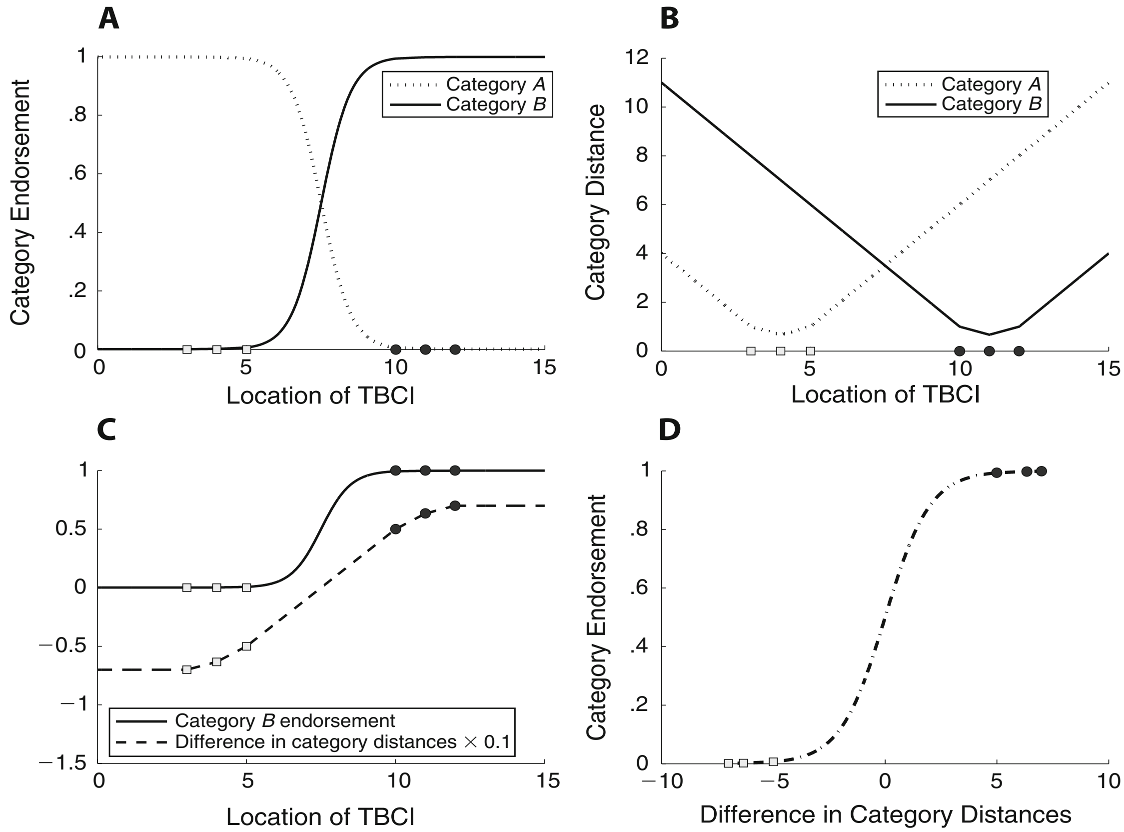

Figure 2. Constructing the distance endorsement relationship for the two categories in Figure 1. (A) Endorsement values for all possible locations of the to-be-classified item (TBCI), calculated by taking the category similarities in Figure 1 and applying the choice rule (Equation 3). The solid line shows the probability of endorsing category *B* (circles), and the dotted line shows the probability of endorsing category *A* (squares). (B) We then calculate the average distance between the TBCI and the category members for all possible locations of the TBCI. Again, the solid lines are category *B*, and the dotted lines are category *A*. (C) The dashed line takes the difference in average distances (i.e., how much closer the TBCI is to category *B* than to category *A*). For comparative purposes, the category endorsement function for category *B* is plotted alongside, again using a solid line. The locations of the various category exemplars on both functions are shown. (D) Plots of the two functions from panel C against one another, using the dash-dotted line, giving us the relationship between distance and endorsement.

cal discussion appears in Luce's (1959, p. 40) treatment of choice probabilities within Fechner's psychophysical theory, for instance. Of course, in Luce's (1959) derivations, the "distance measure" under consideration was rather different from the average measure δ_{iA} used in Equation 5. This provides an illustration of an important point: The derivation of the logistic function depends only on the assumptions that (1) the choice rule applies and (2) it is applied to an exponentiated distance measure. It does not depend on the assumption that the category distance measure is the average interstimulus distance δ_{iA} or a total interstimulus distance $n_A\delta_{iA}$. This article focuses on those measures only because they also appear to be the focus of Smith's (2006) simulations. However, the same approach would work perfectly well for *any* summary measure, not just average distance. For example, we could use the distance from the TBCI to a prototype, which would produce a logistic function that exactly captures the behavior of a prototype model.[4]

The success and failure of the logistic approximation. Readers familiar with the literature on categorization will note that, since the average distance measure δ_{iA} is not actually used by similarity–choice models, the logistic relationship in Equation 5 should not be expected to hold universally. In some cases, a logistic function based on the average distance measure will be a reasonable approximation, but in other cases it will not. For exemplar models in particular, the approximation can be made to fall apart very easily. In such cases, of course, a linear approximation also fails.

However, to begin with, I will provide a simple illustration of cases in which the approximation works. To that end, Figure 3 shows 25 distance–endorsement functions for simple one-dimensional categories that, like the categories in the original example, are constrained not to overlap. Each panel plots a distance endorsement function, in exactly the same way as was done for Figure 2D. However, in this situation, four of the six stimuli were chosen randomly. Specifically, category *A* has one exemplar fixed at 0, and category *B* has one exemplar fixed at 30 (the Appendix discusses the reason for imposing this

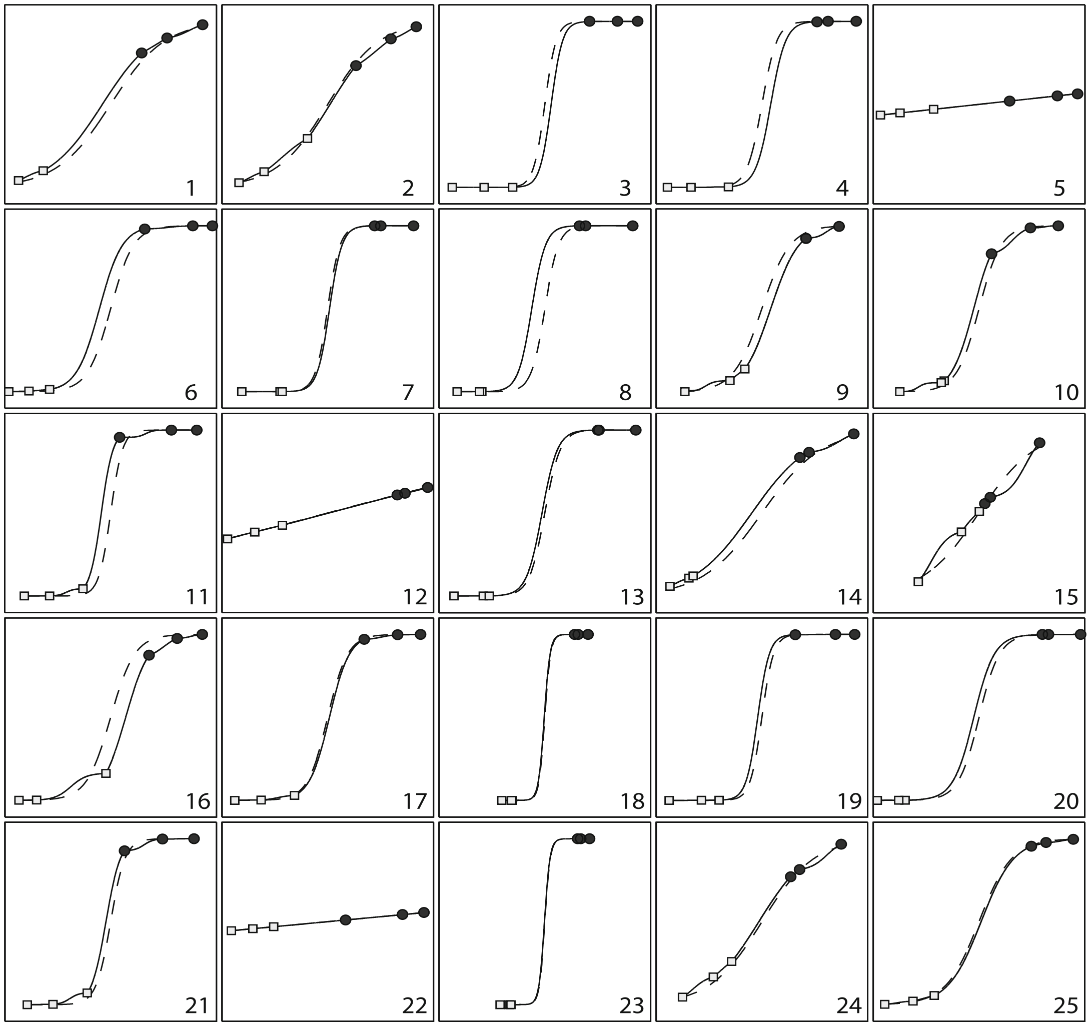

Figure 3. The distance endorsement functions for 25 randomly generated categories. Solid lines depict the exact endorsement function arising from an exemplar model, and the dashed lines represent the logistic approximation. As per the original example in Figure 1, items vary only along a single continuous dimension. Each category has three category members, and the categories are constrained so that they do not overlap. Specifically, category *A* (squares) has one exemplar fixed at 0, and category *B* (circles) has one exemplar fixed at 30. The remaining two exemplars for category *A* are generated from a uniform distribution on the interval [5,15], whereas the remaining two exemplars for category *B* are generated from a uniform distribution on the interval [15,25]. The specificity λ is drawn from a uniform distribution on [0,1].

constraint). The remaining two exemplars for category *A* are generated from a uniform distribution on the interval [5,15], and the remaining two exemplars for category *B* are generated from a uniform distribution on the interval [15,25]. The specificity λ is drawn from a uniform distribution on [0,1]. The true relationships implied by an exemplar model are shown by solid lines, and the approximations are shown by the dashed lines. In some cases, both functions look approximately linear, but this is clearly not true in general. On the other hand, the logistic approximation in Equation 5 is a good fit in all 25 cases (again, the approximation is parameter free, since λ is known).

When we relax the assumption that category similarity and average distance are closely related, the logistic function no longer applies, and the deviations from linearity can become quite dramatic. To give a simple example, Figure 4 presents an example of categories in an exemplar model that yield a very strangely shaped "function," arising because the category similarity functions s_{iA} are not closely related to the exponentiated average distance measure $e^{-\lambda\delta_{iA}}$. In this case, the two categories are interleaved, and as a consequence, the distance–endorsement relationship reverses itself in two places. In fact, we cannot call this a *functional* relationship at all: Formally, a function $y = f(x)$ is not allowed to map a single value of x onto multiple values of y. The solid line in panel C arises by plotting the range of values produced by the function $(x,y) = f(z)$, where z is the location of the TBCI, x is the category distance, and y is the category endorsement. Clearly, the

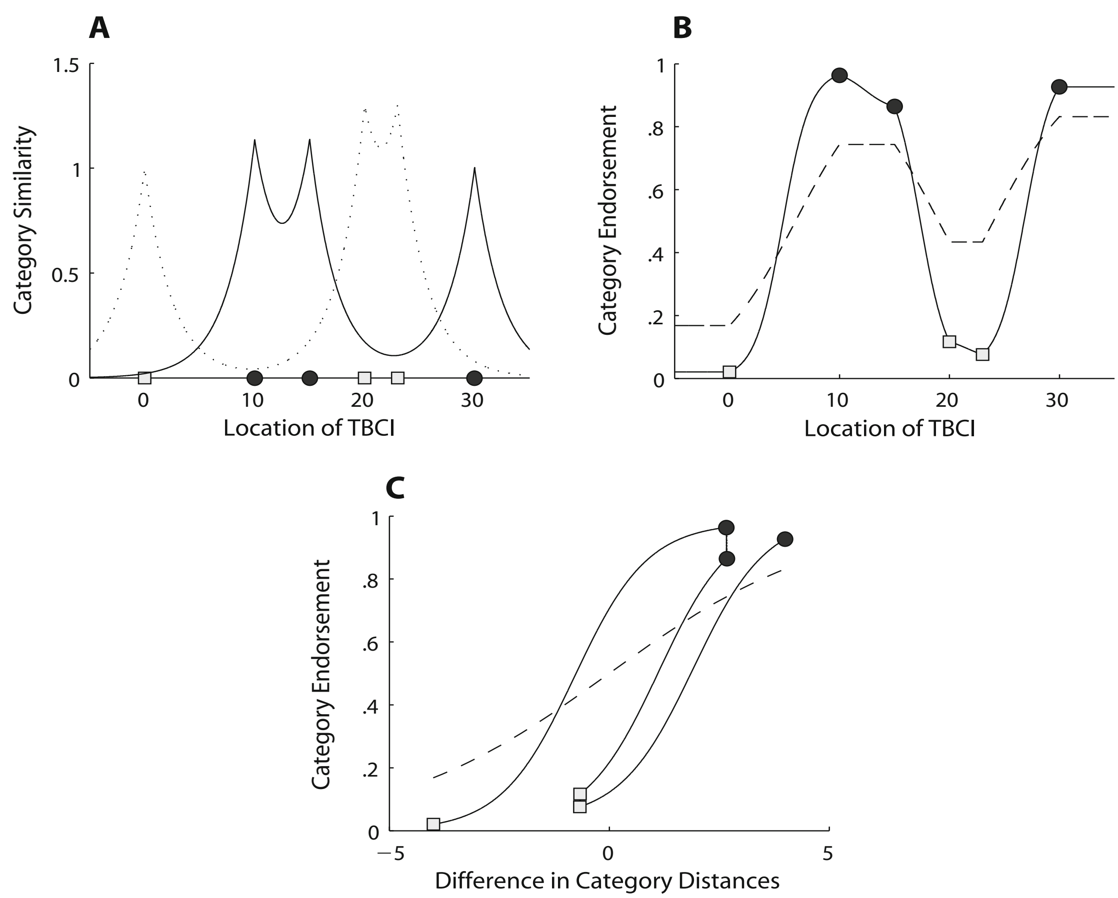

Figure 4. A nonlogistic "function" arising from the use of the average distance measure δ_{iA} in cases in which $e^{-\lambda\delta_{iA}}$ provides a poor approximation to the category similarities s_{iA} in an exemplar model. In this case, the problem arises because the category exemplars are interleaved in a fairly complex fashion (panel A). As a consequence, the logistic approximation behaves poorly (panel B), and the relationship between average distance and category endorsement reverses itself twice (panel C). In panel A, the dotted line shows similarity to category A (squares), and the solid line shows the similarity to category B (circles) in an exemplar model with $\lambda = 0.4$. In panels B and C, the solid lines show the true probability in the model of endorsing each possible TBCI as a member of category B, and the dashed lines show the logistic approximation.

curve that this traces out in the (x,y) space cannot be described by any function of the form $y = f(x)$, making both the linear and the logistic models inappropriate.

An Empirical Example

To summarize the development so far, in those cases in which category similarity is well approximated by an exponential function of a summary measure such as average distance (i.e., when the category structure is "well behaved"), the function relating this distance measure to category endorsement is approximately logistic (not linear) in form. In those cases in which category similarity and average distance dissociate, neither the logistic nor the linear function provides a good approximation in general. This final section will discuss an empirical example involving a well-behaved category structure and will examine the extent to which the logistic and the linear functions can be distinguished in empirical data. Over a small enough range, the logistic function looks approximately linear (as do all smooth functions), so it might very well be the case that the two cannot be easily distinguished by experimental data.

As an attempt to determine the extent to which the two functions can be distinguished, data were taken from the 5 participants in Condition 1 of Experiment 1 reported by McKinley and Nosofsky (1995). Stimuli were circles with a radial line running through them and varied in terms of the angle of the radial line and the size of the circle. Both categories were probabilistic and were defined in terms of a mixture of two bivariate normal distributions. To provide a sense of what the participants experienced, the left panel of Figure 5 plots the stimulus representations for the last 500 items presented to one of the participants, colored by the category that generated the item. In this plot, light squares denote items generated from category A, whereas dark circles denote items generated from category B. The panel on the right shows the same stimuli, but colored by the category endorsed by the participant, rather than by the true generating category. As these plots illustrate, the categories are reasonably well separated and not too com-

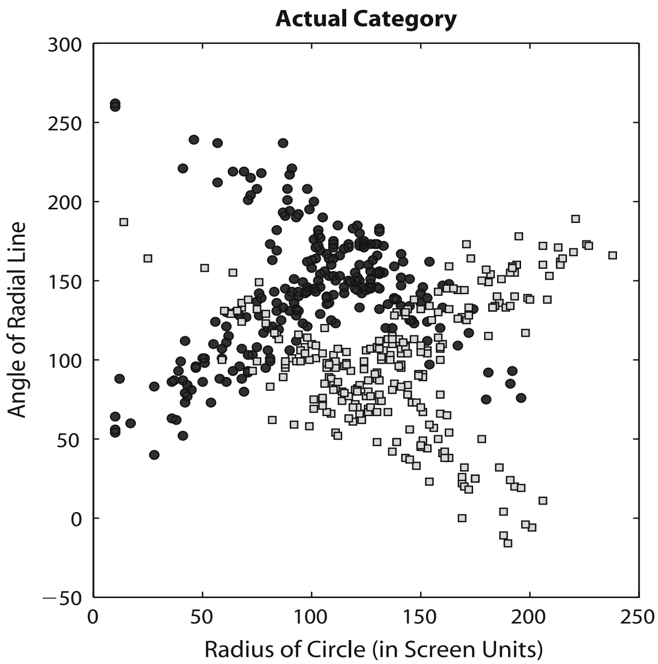

Figure 5. The probabilistic category data for one of the participants (#2) in Experiment 1 of McKinley and Nosofsky (1995), Condition 1. The panel on the left shows the stimulus representation of the last 500 stimuli presented to the participant, colored by the category to which they belong. Stimuli represented using light squares were generated from category *A*, and those represented with dark circles were generated from category *B*. The panel on the right shows the same stimuli, but they are colored by the category to which the participant assigned each stimulus.

plicated in their shape. As a consequence, there is some reason to suspect that average distance and category similarity will not strongly dissociate in this task and that the logistic approximation will be effective.

In order to construct the empirical distance endorsement function for each of the 5 participants, we can calculate the average distance from each of the 4,000 or so observed stimuli to the two categories. These distances are then binned (e.g., the first bin includes all stimuli that lie 60–70 units closer to category *B* than to category *A*), so that we can calculate, for each bin, the probability that each person would choose category *A*. The resulting distance endorsement relationships for each of the 5 participants are plotted in Figure 6. The error bars are 95% confidence intervals, calculated independently for each participant and each bin. Simple visual inspection suggests that all five functions are curvilinear. To see how well the two models perform, I fit a linear function of the form $y = \frac{1}{2} + \beta x$ and a logistic function of the form $y = (1 + e^{-\lambda x})^{-1}$, where x denotes the difference in average distance and y denotes category endorsement. Both functions have one free parameter, and so, in a simple analysis, we might evaluate the two by choosing β and λ to maximize the variance accounted for (i.e., r^2) by the two models. Since the five individual data sets are almost identical, it is reasonable to average them (for more formal discussions, see Lee & Webb, 2005; Navarro, Griffiths, Steyvers, & Lee, 2006) before fitting the two models. In doing so, it turns out that the logistic function can explain 99.0% of the variance, whereas a linear function explains only 93.6% of the variance.[5] To a close approximation, the predicted logistic relationship holds in this case.

Discussion

The analysis of the internal structure of models can be extremely useful. In one particularly impressive case, an analysis of the geometry of neural network state spaces (see Amari, Park, & Fukumizu, 2000) demonstrated that "plateaus" in the learning process, not unusual in connectionist models, often arise because the standard backpropagation learning rule actually distorts the state space, creating "singularities" that slow down the learning in particular regions. Since it is unlikely that these singularities have any genuine psychological equivalent, it is important to be aware of this structural problem. In general, identifying pathologies of this kind is important for the advancement of modeling practice. To the extent that the article by Smith (2006) raises awareness of these issues as they pertain to categorization models, it performs a valuable function.

However, when the behavior of a model is analyzed, it is important to avoid painting a misleading picture. In the case of similarity–choice models for categorization, the function relating distance to category membership is not linear. Rather, under those conditions in which category similarity is approximately exponentially related to an appropriate category distance measure, the typical form of the predicted relationship is logistic. Obviously, not all domains are able to discriminate between the logistic function and a linear one, a fact that Smith's (2006) simu-

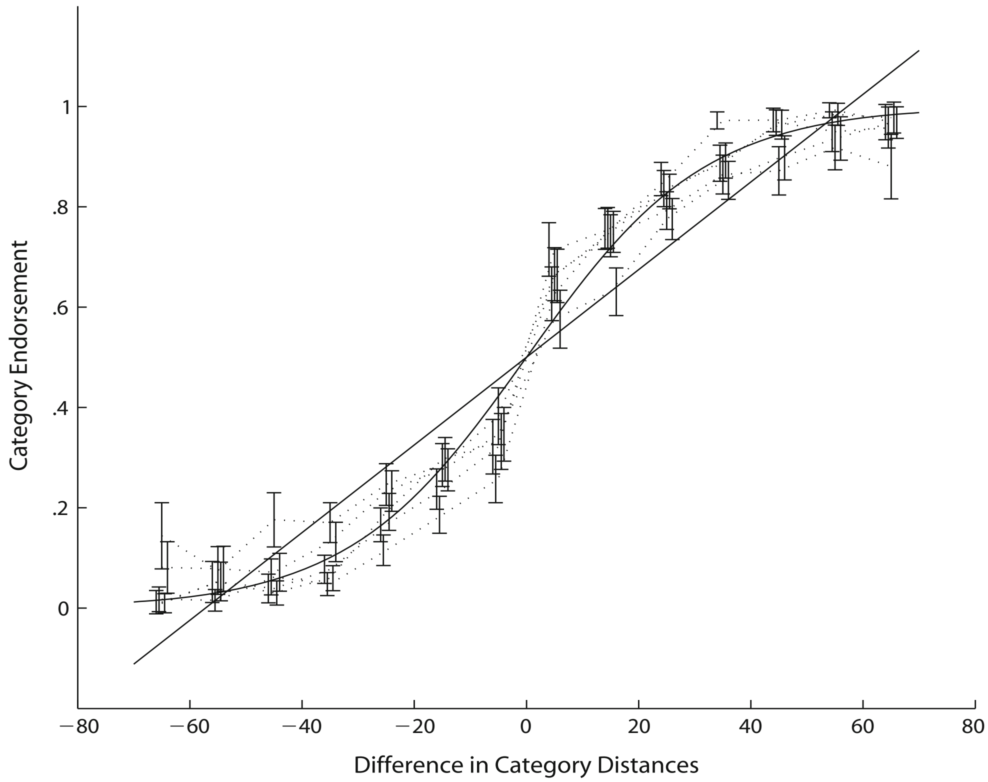

Figure 6. Response probabilities versus average distance discrepancy for each of 5 participants in Experiment 1 of McKinley and Nosofsky (1995), Condition 1. The five empirical functions (dotted lines) were constructed by binning the raw stimuli (see the main text for details), with error bars representing 95% confidence intervals for each bin. The two solid lines show the best-fitting linear and logistic models; the linear function explains 93.6% of the variance, whereas the logistic function explains 99.0% of the variance.

lations illustrate. This should not, however, be taken as evidence that the general form of the relationship is linear; there are some domains (and corresponding empirical data sets) that can and do show that the functions are curvilinear (e.g., McKinley & Nosofsky, 1995) and well approximated by a logistic equation. More generally, however, it would be equally incorrect to assume that the relationship is invariably logistic. The derivation of a logistic relationship relies on the assumption that category distance and category similarity do not strongly dissociate. When they do, as is often the case when an average distance measure is applied to highly overlapping categories, the relationship can become extremely pathological, and it would be a mistake to apply *any* function of category distance.

As a final remark, it is worth noting that it would be odd *not* to get some kind of sigmoidal relationship when the categories are "well behaved." For instance, it is generally accepted that logistic models are statistically superior to linear models for the analysis and interpretation of categorical data (Hosmer & Lemeshow, 2000); so, to the extent that human categorization can be regarded as a rational solution of a statistical problem (e.g., Ashby & Alfonso-Reese, 1995), we should expect to see a performance profile of this form. At a more empirical level, there are strong constraints implied by the psychophysics literature. Consider the special case in which people are given one item from each class and these two items, x_1 and x_2, differ only on a single perceptual dimension (e.g., length). Each item would be considered to be the sole representative of (or standard for) two perceptually based category responses. In our hypothetical experiment, we would then present people with a third item, x_3, that lies somewhere between the two and would ask participants to decide whether it should be classified with the first or the second item. In essence, we would be asking people to judge whether the perceptual difference $|x_3 - x_1|$ is greater than or less than the perceptual difference $|x_3 - x_2|$. When phrased in these terms, it is clear that this task is very similar to psychophysical experiments for measuring discrimination thresholds via a two-alternative forced choice method (e.g., Woodsworth & Schlosberg, 1954, chap. 20). Although there are a number of technical issues regarding how to estimate the psychometric functions that emerge from this kind of procedure (e.g., Wichmann & Hill, 2001) and some questions over their precise shape, few would disagree with the assertion that psychometric functions are generally nonlinear and approximately sigmoidal. If similarity–choice models predicted a linear endorsement function for a psychophysics-style experiment, they would not be taken seriously as models for human behavior. Fortunately, it is trivial to show that in this case, in which only a single instance of each category is known, the relationship $s_{iA} = e^{-\lambda\delta_{iA}}$ holds exactly, so the function is logistic irrespective of whether an exemplar model or a prototype model is used.

AUTHOR NOTE

This research was supported by Australian Research Fellowship Grant DP-0773794. I thank Yves Rosseel for providing the data used in this article, as well as Nancy Briggs, Anna Ma-Wyatt, and the reviewers and action editor for helpful comments. Correspondence concerning this article should be addressed to D. J. Navarro, School of Psychology, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: daniel .navarro@adelaide.edu.au).

REFERENCES

AMARI, S.-I., PARK, H., & FUKUMIZU, K. (2000). Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural Computation*, **12**, 1399-1409.

ASHBY, F. G., & ALFONSO-REESE, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, **39**, 216-233.

BALASUBRAMANIAN, V. (1997). Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. *Neural Computation*, **9**, 349-368.

BROWNE, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, **44**, 108-132.

GOODMAN, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437-447). New York: Bobbs-Merrill.

HOSMER, D. W., & LEMESHOW, S. (2000). *Applied logistic regression*. New York: Wiley.

LEE, M. D., & WEBB, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, **12**, 605-621.

LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.

LUCE, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, **15**, 215-233.

MCKINLEY, S. C., & NOSOFSKY, R. M. (1995). Investigations of exemplar and decision-bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

MYUNG, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, **44**, 190-204.

MYUNG, I. J., KIM, C., & PITT, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition*, **28**, 832-840.

MYUNG, I. J., & PITT, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, **4**, 79-95.

NAVARRO, D. J. (2005). Analyzing the RULEX model of category learning. *Journal of Mathematical Psychology*, **49**, 259-275.

NAVARRO, D. J. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, **51**, 85-98.

NAVARRO, D. J., GRIFFITHS, T. L., STEYVERS, M., & LEE, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, **50**, 101-122.

NAVARRO, D. J., PITT, M. A., & MYUNG, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, **49**, 47-84.

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

PITT, M. A., KIM, W., NAVARRO, D. J., & MYUNG, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, **113**, 57-83.

PITT, M. A., MYUNG, I. J., & ZHANG, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.

ROBERTS, S., & PASHLER, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, **107**, 358-367.

ROSCH, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-77). Hillsdale, NJ: Erlbaum.

SHEPARD, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, **210**, 390-398.

SHEPARD, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

SMITH, J. D. (2006). When parameters collide: A warning about categorization models. *Psychonomic Bulletin & Review*, **13**, 743-751.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B*, **39**, 44-47.

THOMPSON, A. C. (1996). *Minkowski geometry*. Cambridge: Cambridge University Press.

TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

WICHMANN, F. A., & HILL, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293-1313.

WOODSWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology* (3rd ed.). London: Methuen.

NOTES

1. The notion is that the category similarities s_{iA} are assumed to act as response strengths $v(A)$ in the sense discussed by Luce (1959, Theorem 3, p. 23). However, since similarity–choice models aim to specify this scale directly (and by assumption), we should perhaps say that the adoption of this rule is "loosely consistent" with Luce's choice axiom. Strictly, we should note that the choice axiom merely implies the existence of *some* positive real-valued function $v(\cdot)$ that is defined for all responses that have nonzero probability, so that the function is independent of the choice set and this normalization rule holds (see Luce, 1977 for an overview).

2. When calculating these correlations, I assumed that each location for the TBCI in the range [0,15] was equally likely, as per Figure 1. Note that a slightly different result would be obtained if we assumed that each difference in category distance (as per Figure 2D) was equally likely.

3. The measure used in Simulation 4 (and hence, Simulations 6 and 10) is not entirely clear from the description in the article, which provides a list of interstimulus distances but does not explicitly state what category distance measure was used. However, the article does indicate that Simulation 4 is analogous to Simulation 2, so it appears that an average measure was used in Simulations 4, 6, and 10. In Simulation 8, the distance to category was not inferred from any stimulus representation but was, instead, directly specified.

4. As an alternative, we could propose a "distance" measure of the form $(1/\lambda)\ln \Sigma_{j\in A} e^{-\lambda d_{ij}}$, which would yield a logistic function that exactly mimics an exemplar model, since this measure is equivalent to $(1/\lambda)\ln s_{iA}$. However, without a good a priori reason to use this as a genuine measure of category distance, it would be very misleading to talk about an exact logistic distance endorsement function for the exemplar model.

5. Readers familiar with the model selection literature will recognize that a simple r^2 measure of this kind is not always safe, since superior performance can be an artifact of model complexity (Myung, 2000). To provide a simple check that this is not the case here, I also undertook a cross-validation analysis (e.g., Browne, 2000; Stone, 1974), using 10,000 random 50–50 splits of the data into training sets and test sets. For the logistic function, the median value for the cross-validated fit remained very high, at 99% of the variance in the test set. For the linear function, the median fit fell slightly, to 92% of the variance. More important, in approximately 93% of the random splits, cross-validation prefers the logistic model.

(*Continued on next page*)

APPENDIX

In this appendix, I present a somewhat more precise discussion of the logistic approximation. The main text implicitly assumes an unweighted city block metric for measuring interstimulus distances—not unreasonably, since stimulus dimensions are assumed to be separable—and this metric is a special case of both the Minkowski metrics (see Thompson, 1996, chap. 1), which are typically used with continuous dimensions, and Tversky's (1977) contrast model, which is more often applied for binary features. This is useful since it is one of the few cases in which geometric models and featural models make the same assumptions about distance and dissimilarity. Accordingly, the assumption of an unweighted city block metric will be retained for the following derivation, although very few elements of the derivation rely on it and it would not be too difficult to repeat the exercise for different distance measures. For clarity of expression, in what follows, j will always index a category exemplar, and k will always index a dimension, so the summation notation will be shortened to Σ_j and Σ_k. I will also switch to the exp( ) notation for exponentiation. Taking Shepard's (1987) generalization law as a primitive, the construction of category similarities follows the following rule for a prototype model:

$$s_{iA} = \exp\left(-\lambda d_{i\bar{A}}\right), \tag{A1}$$

where $d_{i\bar{A}}$ denotes the distance from the prototype to the TBCI, which for the unweighted city block metric is simply

$$\begin{aligned} d_{i\bar{A}} &= \sum_k \left| \left(1/n_A\right)\left(\sum\nolimits_j x_{jk}\right) - x_{ik} \right| \\ &= \left(1/n_A\right)\sum_k \left| \sum\nolimits_j x_{jk} - x_{ik} \right|. \end{aligned} \tag{A2}$$

For an exemplar model, the expression is somewhat different:

$$s_{iA} = \sum_j s_{ij} = \sum_j \exp\left(-\lambda d_{ij}\right). \tag{A3}$$

Recall that the average distance measure δ_{iA} can be expressed as

$$\delta_{iA} = \left(1/n_A\right)\sum_j d_{ij}. \tag{A4}$$

At this point, we can begin a somewhat more careful construction of the logistic approximation. For both exemplar and prototype models, the starting point is to rewrite the category similarity function s_{iA} in a more tractable form. For a prototype model,

$$\begin{aligned} s_{iA} &= \exp\left(-\lambda\left[\delta_{iA} - \delta_{iA} + d_{i\bar{A}}\right]\right) \\ &= \exp\left(-\lambda\delta_{iA} - q_{iA}\right), \end{aligned} \tag{A5}$$

where we define

$$q_{iA} = \lambda\left[d_{i\bar{A}} - \delta_{iA}\right]. \tag{A6}$$

The exemplar model can be rewritten in much the same way, but in this case, the expression is given by

$$\begin{aligned} s_{iA} &= \exp\left(-\lambda\delta_{iA} + \left[\ln\sum_j \exp\left(-\lambda d_{ij}\right)\right] + \lambda\delta_{iA}\right) \\ &= \exp\left(-\lambda\delta_{iA} - q_{iA}\right), \end{aligned} \tag{A7}$$

where q_{iA} in this case becomes

$$q_{iA} = -\left[\ln\sum_j \exp\left(-\lambda d_{ij}\right)\right] - \lambda\delta_{iA}. \tag{A8}$$

Note that the term in the square brackets here is simply the logarithm of the category similarity (i.e., s_{iA}), where the category similarity is written as a function of the interstimulus distances, rather than the more typical expression based on interstimulus similarities. This way of writing things is to be preferred, since the goal is to discuss the direct relationship between distances and category endorsement. In any case, observe that $q_{iA} = -\ln s_{iA} - \lambda\delta_{iA}$ for both the prototype and the exemplar models, representing the extent to which category similarity and category distance dissociate.

The reason for rewriting both models in this fashion is to allow the category endorsement $P(i \in A)$ to be expressed as a function of the distances δ_{iA} and δ_{iB}, as well as q_{iA} and q_{iB}:

APPENDIX (Continued)

$$P(i \in A) = \frac{s_{iA}}{s_{iA} + s_{iB}}$$
$$= \frac{\exp(-\lambda\delta_{iA} - q_{iA})}{\exp(-\lambda\delta_{iA} - q_{iA}) + \exp(-\lambda\delta_{iB} - q_{iB})}$$
$$= \frac{1}{1 + \exp(-\lambda[\delta_{iB} - \delta_{iA}] - [q_{iB} - q_{iA}])}. \tag{A9}$$

We arrive at the logistic approximation by assuming that $q_{iB} - q_{iA} \approx 0$. There are a number of special cases for which the approximation holds. The obvious one, discussed in the main text, is when $s_{iA} \approx \exp(-\lambda\delta_{iA})$ (for both categories), since this implies that $q_{iA} \approx 0$ and $q_{iB} \approx 0$ and, therefore, $q_{iB} - q_{iA} \approx 0$.

More generally, it is relatively straightforward to derive an informative expression for the error associated with the approximation. To do so, we define $u = \exp(-\lambda[\delta_{\mathrm{iB}} - \delta_{iA}]/2)$ and $v = \exp(-[q_{iB} - q_{iA}]/2)$. If we let $P^*(i \in A)$ denote the approximated endorsement level, the error of approximation is given by

$$P(i \in A) - P^*(i \in A) = \frac{1}{1+u^2v^2} - \frac{1}{1+u^2}$$
$$= \frac{u}{u^2+1}\frac{u^2+1}{u}\left[\frac{1}{1+u^2v^2} - \frac{1}{1+u^2}\right]$$
$$= \frac{u}{u^2+1}\left[\frac{uv^2-u}{u^2v^2+1}\right]$$
$$= \frac{u}{u^2+1}\left[\frac{v-v^{-1}}{uv+u^{-1}v^{-1}}\right]. \tag{A10}$$

Writing the error in this form is useful, since

$$\frac{u}{u^2+1} = \frac{1}{\sqrt{u^2+1}}\frac{u}{\sqrt{u^2+1}}$$
$$= \sqrt{\frac{1}{u^2+1}\frac{u^2}{u^2+1}}$$
$$= \sqrt{\frac{1}{u^2+1}\left[1 - \frac{1}{u^2+1}\right]}$$
$$= \sqrt{P^*(i \in A)\left[1 - P^*(i \in A)\right]}. \tag{A11}$$

This expression takes its maximum value of 0.5 when $P^*(i \in A) = 1 - P^*(i \in A) = 0.5$ and diminishes quickly as $P^*(i \in A) \to 0$ or $P^*(i \in A) \to 1$. This term acts to reduce the error at either end of the curve, although this will not necessarily be successful in pathological cases. However, throughout this article, I have focused mainly on categories that are at least partially separable, in the sense that the category with the leftmost mean (in the unidimensional case) has at least one exemplar that is substantially further to the left than any of the exemplars from the opposing category (and the reverse is true for the rightmost category). By imposing this constraint, both the true function and the logistic approximation will usually (but not always, especially if λ is very small) involve cases in which $P^*(i \in A) \to 0$ or $P^*(i \in A) \to 1$. When this constraint is removed, larger errors are observed.

Turning to the second term in the expression, we see that the other major determinant of error is the extent to which distance and similarity dissociate, since

$$\frac{v-v^{-1}}{uv+u^{-1}v^{-1}} = \frac{\exp(-[q_{iB} - q_{iA}]/2) - \exp([q_{iB} - q_{iA}]/2)}{\exp(-\lambda[(\delta_{iB} - \delta_{iA}) + (q_{iB} - q_{iA})]/2) + \exp(\lambda[(\delta_{iB} - \delta_{iA}) + (q_{iB} - q_{iA})]/2)}. \tag{A12}$$

It is this term that describes how the approximation breaks down when the category distance becomes a poorer approximation to the log similarities. As $|q_{iB} - q_{iA}|$ becomes large, so too does this term. Although it would presumably be possible to further analyze this term and find deeper descriptions of those characteristics that lead to violations of the logistic approximation, any such analysis is beyond the scope of this article.

(Manuscript received January 27, 2007;
revision accepted for publication May 11, 2007.)