Memory & Cognition
2008, 36 (6), 1057-1065
doi: 10.3758/MC.36.6.1057

Putting the psychology back into psychological models: Mechanistic versus rational approaches

Yasuaki Sakamoto
Stevens Institute of Technology, Hoboken, New Jersey

Matt Jones
University of Colorado, Boulder, Colorado

and

Bradley C. Love
University of Texas, Austin, Texas

Two basic approaches to explaining the nature of the mind are the rational and the mechanistic approaches. Rational analyses attempt to characterize the environment and the behavioral outcomes that humans seek to optimize, whereas mechanistic models attempt to simulate human behavior using processes and representations analogous to those used by humans. We compared these approaches with regard to their accounts of how humans learn the variability of categories. The mechanistic model departs in subtle ways from rational principles. In particular, the mechanistic model incrementally updates its estimates of category means and variances through error-driven learning, based on discrepancies between new category members and the current representation of each category. The model yields a prediction, which we verify, regarding the effects of order manipulations that the rational approach does not anticipate. Although both rational and mechanistic models can successfully postdict known findings, we suggest that psychological advances are driven primarily by consideration of process and representation and that rational accounts trail these breakthroughs.

Two basic approaches to explaining the nature of the mind are the rational and the mechanistic approaches. Rational analyses attempt to characterize the environment and the behavioral outcomes that humans seek to optimize. The rational approach holds that people are adaptive and learn (at the individual or species level) to behave optimally given the nature of the environment (i.e., given available information or statistics). The formal product of a rational analysis is an abstract mathematical model (often Bayesian) that details the behavioral strategies that optimize some cost function, given the environment. Such models do not have recourse to how people actually process and represent information but are, instead, abstract.

Considerations of the environment and optimality also resonate with adherents of the mechanistic program, but unlike for a rational model, the main goal of a mechanistic model is to simulate human behavior by using mechanisms (i.e., analogous processes and representations) that are the same as those that support human behavior. The mechanistic program seeks to reverse engineer the human brain and peer inside the black box. The issues of primary importance to the mechanistic program are how people represent and process information.

One common criticism of mechanistic approaches is that they lead to ad hoc explanations that lack the elegance and clarity of models derived from rational analysis. To the extent that two models converge on a common set of predictions, the more transparent and mathematically motivated model should be favored. Echoing these sentiments, Anderson (1991b) stated, “All mechanistic proposals which implement the same rational prescription are the same,” and “a rational theory provides a precise characterization and justification of the behavior the mechanistic theory should achieve.” These views are seconded by Chater and Oaksford (1999): “The picture that emerges from this focus on mechanistic explanation is of the cognitive system as an assortment of apparently arbitrary mechanisms, subject to equally capricious limitations, with no apparent rationale or purpose.”

The upshot of these statements is that mechanisms are subservient to rational accounts of thought. Perhaps in a moment of candor or euphoria, Anderson (1991b) stated that rational models render mechanistic models unnecessary: “One might take the view (and I have so argued in overenthusiastic moments; Anderson, in press) that we do not need a mechanistic theory, that a rational theory offers a more appropriate explanatory level for behavioral

B. C. Love, brad_love@mail.utexas.edu

data" (p. 471). We believe that these general sentiments explain the rising popularity of rational accounts of cognition (for reviews, see Chater, Tenenbaum, & Yuille, 2006; Griffiths, Kemp, & Tenenbaum, 2008).

In this article, we advance a different view of mechanistic and rational models. Rather than viewing the details of mechanistic models as arbitrary, we argue that these differences are key to generating novel predictions. We do not view mechanistic models as simply implementing rational models. We argue, by way of demonstration, that mechanistic and rational models are likely to diverge in important ways once the full entailments of the mechanistic model are appreciated. In other words, mechanistic models can motivate predictions beyond those of a successful rational analysis. Thus, mechanistic models are properly understood as driving theory advancement, rather than as bastardized instantiations of more abstract rational analyses.

Of course, both rational and mechanistic accounts can be constructed after the fact to account for any data set. These two approaches can also be viewed as complementary and compatible in that they address behavioral phenomena at different levels of explanation (see Marr, 1982). To be clear, our key metascientific argument is that mechanistic models are best suited for deriving surprising behavioral predictions.

To offer tentative support for our metascientific argument and to make an independent empirical contribution, we conducted two experiments in which we examined how people learn about the variance of categories. Experiment 1 was in the tradition of studies exploring people's sensitivity to category variability (e.g., Cohen, Nosofsky, & Zaki, 2001; Fried & Holyoak, 1984; Hahn, Bailey, & Elvin, 2005). Experiment 2 expanded on Experiment 1 to consider how trial order impacts perceptions of category variability. The domain we chose was a simple category-learning task in which mechanistic and rational accounts were already fleshed out.

An initial experiment in which the role of category variability in generalization was explored suggested obvious models from within both perspectives. The two models were largely in accord, but an examination of how the mechanistic model built internal representations of the categories in response to corrective feedback suggested a second experiment in which the predictions of the two accounts diverged and for which the results supported the mechanistic account. Empirically, in Experiment 2, we demonstrated how people's impressions of category variability could be strongly affected by manipulating the order in which category members were experienced.

The unique predictions of the mechanistic model followed from the insight that people incrementally build representations in memory, rather than from any insight into the structure of the environment. In effect, the mechanistic model suggested revision of the rational account—a direction of theory development opposite that advocated by proponents of rational analysis.

EXPERIMENT 1

Fried and Holyoak (1984) found that after training on two contrasting categories of unequal variance, subjects tended to classify intermediate items into the higher variance category. This sensitivity to category variance was verified in subsequent learning studies (e.g., Cohen et al., 2001; Hahn et al., 2005). Preferences in generalizing to high-dispersion categories have also been found in experiments that tapped preexisting knowledge and categories (Rips, 1989), as opposed to utilizing learning procedures.

Experiment 1 refined aspects of previous learning studies. In Experiment 1, subjects learned to classify lines varying in length into one of two categories. The design is illustrated in Figure 1. Learning items are illustrated as dark triangles. The six items (L1–L6) forming one category are less variable than the six items (H1–H6) forming the contrasting category. Following learning, the subjects classified a variety of items, including some items that were not experienced during learning, such as Item N6. These novel items were tests of how subjects generalize. Item N6 was of particular interest since it was midway between the nearest trained members (L6 and H1) of the low- and high-dispersion categories.

To foreshadow, our results replicated previous findings indicating that people generalize border items to the high-dispersion category. After the method and results have been presented, mechanistic and rational models will be derived and fit to the data.

Method

Fifty University of Texas undergraduates learned to correctly assign 12 line stimuli (represented by dark triangles labeled L1–L6 and H1–H6 in Figure 1) into Category A or B through trial-by-trial classification learning with corrective feedback. The members of one category (L1–L6) varied relatively little in their lengths, whereas the members of the other category (H1–H6) were highly variable. The stimulus lengths in pixels (100 pixels = 33.25 mm) are presented in Figure 1. To eliminate possible influences of absolute line length on performance (Ono, 1967), whether the high-dispersion category had longer or shorter lines than the low-dispersion category was counterbalanced between subjects (see Figure 1). The border item (N6) had the same length in both conditions.

On each training trial, one line was presented horizontally at the center of a display, and the text "Category A or B?" appeared at the top left corner of the display. After responding A or B, the subjects received visual (e.g., "Right! The correct answer is A." or "Wrong! The correct answer is B.") and auditory (a low-pitch tone for errors and a high-pitch tone for correct responses) corrective feedback. The visual feedback (presented at the bottom left corner of the display) and the stimulus were displayed for 2,000 msec after the subjects had responded. The subjects completed 10 blocks of training trials. A block comprised presentation of every training item in a random order. The density curves shown in Figure 1 are illustrative of possible mental representations, as discussed below, and do not indicate information about the frequency of presentation during the experiment.

Following training, the subjects answered three addition problems to prevent rehearsal of information from the learning phase. Finally, the subjects completed two blocks of transfer classification. In each transfer block, the subjects classified the 12 studied items and 11 novel items (represented by light triangles labeled N1–N11 in Figure 1) in a random order as they did in the training phase, except that no corrective feedback was provided in the transfer phase. Our main interest was the subjects' performance on the border transfer item (N6) that was midway between the nearest studied members (L6 and H1) of the two categories.

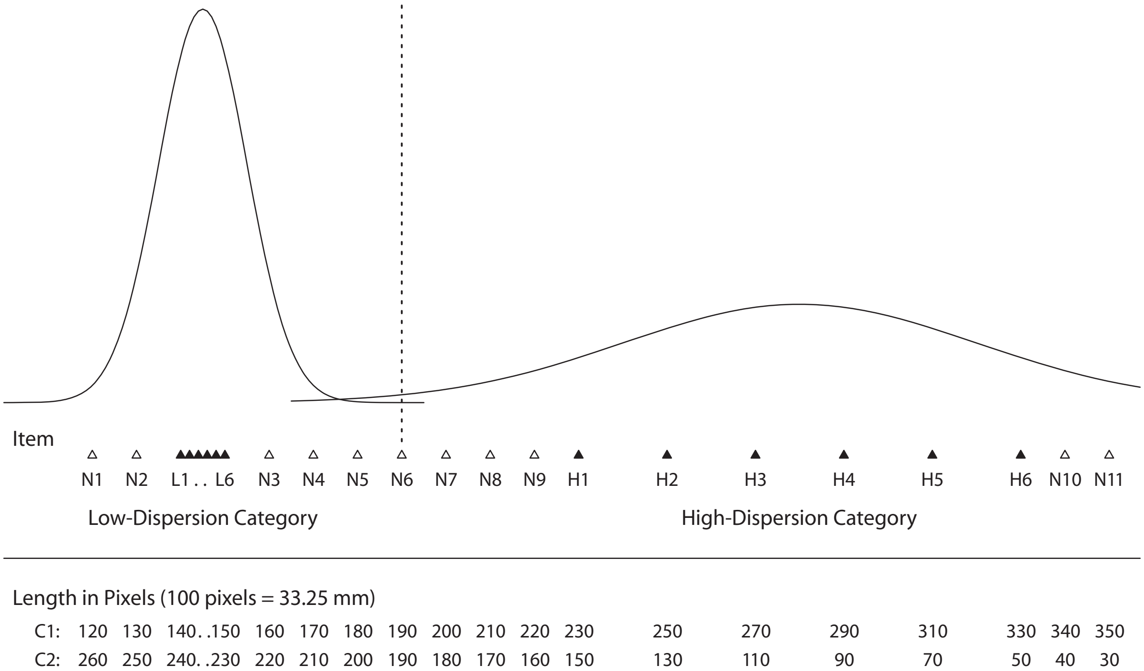
Item
N1 N2 L1 . . L6 N3 N4 N5 N6 N7 N8 N9 H1 H2 H3 H4 H5 H6 N10 N11
Low-Dispersion Category
High-Dispersion Category
Length in Pixels (100 pixels = 33.25 mm)
C1: 120 130 140. .150 160 170 180 190 200 210 220 230 250 270 290 310 330 340 350
C2: 260 250 240. .230 220 210 200 190 180 170 160 150 130 110 90 70 50 40 30

Figure 1. The design of Experiment 1. Dark triangles (L1–L6 and H1–H6) represent studied items, and light triangles (N1–N11) represent novel items that did not appear during learning. The item lengths are spaced to scale. Item N6 is exactly midway between the nearest studied members (L6 and H1) of the low- and high-dispersion categories. Each studied item in the low-dispersion category differs from its nearest neighbor by 2 pixels, whereas each studied item in the high-dispersion category differs from its nearest neighbor by 20 pixels. To eliminate possible influences of absolute line length on performance, whether the high-dispersion category had longer (Condition C1) or shorter (Condition C2) lines than the low-dispersion category was counterbalanced between subjects. The two density functions are illustrative of the category representations developed by both rational and mechanistic models when applied to this task.

Results

Border Item N6 was more likely to be classified into the high- than into the low-dispersion category. As is shown in Figure 2, averaged across the two transfer blocks, the subjects assigned the border item to the high-dispersion category with greater-than-chance probability [.69 vs. .5; $t(49) = 3.86$, $p < .001$]. In the first transfer block, more subjects (33 of 50) classified the border item into the high-dispersion category than was expected by chance (exact binomial $p = .033$, two-tailed). The same pattern (36 of 50) was found for Item N6 in the second transfer block (exact binomial $p = .0026$, two-tailed).

Rational and Mechanistic Models

Straightforward rational analyses, whether following a maximum likelihood (e.g., Fried & Holyoak, 1984) or a Bayesian (e.g., Tenenbaum & Griffiths, 2001) canon, converge in their account of Experiment 1. A rational analysis of Experiment 1 suggests a model that estimates the true mean and variance of each category on the basis of the unbiased integration of information conveyed by the training items. Although these estimates can be made incrementally (e.g., the current trial's posterior distribution serves as the next trial's prior distribution in a Bayesian scheme), they are equivalent to estimating the mean and variance on the basis of all experienced items (i.e., perfect and unbiased memory). From these estimated means and variances, the probability that a novel item belongs to each category can be calculated, and the item can be assigned to the more likely category. One such model is the unequal variance signal detection model (Green & Swets, 1966; Maddox & Ashby, 1998) when the standard deviation and mean of each category distribution are estimated from all previous learning trials. These rational models correctly predict that Border Item N6 will be assigned to the high-dispersion category.

To facilitate comparison, we derive a mechanistic model that principally differs from the aforementioned rational models in that the mechanistic model does not have perfect memory for the training items. Instead, it sequentially updates its representation of each category (both mean and dispersion) on the basis of the current stimulus, using error-driven learning. Like the rational models, the mechanistic model represents each category in terms of its mean and variance. This information is represented by a cluster for that category (cf. Anderson, 1991a). The cluster tracks the prototype of the category while also encoding its variability. This model is more correctly viewed as a suitable comparison with the aforementioned rational models and as a distillation and simplification of previous proposals than as a new model. Related mechanistic proposals have extended multiple prototype models (Love & Jones, 2006, which extends Love, Medin, & Gureckis, 2004) and ex-

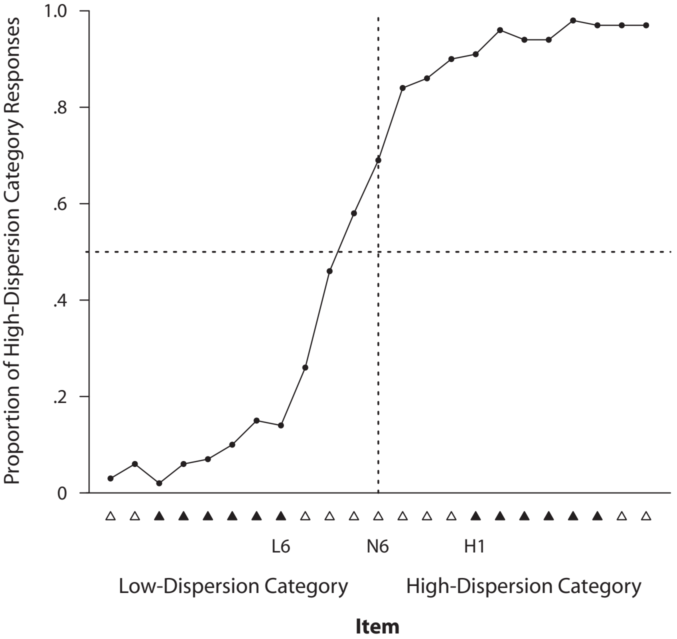

Figure 2. The probability of subjects' classifying each stimulus item as a member of the high-dispersion category during the transfer phase of Experiment 1. Training items are shown as dark triangles; novel items are shown as light triangles. Item N6 is midway between the nearest studied members (L6 and H1) of the low- and high-dispersion categories. Items are not spaced to scale (see Figure 1 for the physical scale).

emplar models (Rodrigues & Murre, 2007, and Sakamoto, Matsuka, & Love, 2004, extend Kruschke, 1992).[1]

Activation of cluster i, a_i, represents the strength of evidence that a stimulus belongs to category i and is a Gaussian function of the presented stimulus value, x:

$$a_i = \frac{1}{\sqrt{2\pi}\,s_i} e^{-\frac{(x-m_i)^2}{2s_i^2}}, \tag{1}$$

where m_i and s_i are the cluster's mean and standard deviation, respectively. The generalization gradient of a category is captured by s_i. The response probability for each category is proportional to the activation of the corresponding cluster (i.e., the probability matching response rule).

Cluster means and standard deviations are updated by gradient descent on an error, $E = ½(t_i - a_i)^2$:

$$\Delta m_i = -\varepsilon_m \frac{\partial E}{\partial m_i} = \varepsilon_m (t_i - a_i) \frac{x - m_i}{s_i^3 \sqrt{2\pi}} e^{-\frac{(x-m_i)^2}{2s_i^2}} \tag{2}$$

and

$$\Delta s_i = -\varepsilon_s \frac{\partial E}{\partial s_i} = \varepsilon_s (t_i - a_i) \frac{(x - m_i)^2 - s_i^2}{s_i^4 \sqrt{2\pi}} e^{-\frac{(x-m_i)^2}{2s_i^2}}, \tag{3}$$

where ε_m and ε_s are learning rates for cluster means and standard deviations, respectively, and t_i is the feedback to cluster i, equal to α if the stimulus is in category i and to 0 otherwise. Cluster means are initialized at the value of the first presented stimulus in each category, and standard deviations are initialized at s_0.

The mechanistic model was trained and tested in a trial-by-trial fashion paralleling the procedure used with the human subjects. Figure 3 illustrates the dynamics of the model simulated on Experiment 1. This figure is based on an average over 10,000 separate runs, using the parameter values $s_0 = 20$, $\alpha = .05$, $\varepsilon_s = 70{,}000$, and $\varepsilon_m = 98{,}000$. These parameters were chosen to fit data from Experiments 1 and 2 simultaneously, but the qualitative results of both experiments were robust to the majority of the parameter space explored.

Prior to training, both clusters have the same standard deviation of 20, and Border Item N6 is closer to the cluster representing the low-variability category. Thus, Item N6 should initially be assigned to the low-variability category, since it more strongly activates that category's cluster. To confirm this intuition, 25 University of Texas undergraduates were shown the two category prototypes (no other training) and chose the category to which the border stimulus belonged. In this single triad task, 22 of 25 subjects preferred to classify Border Item N6 into the low-dispersion category (i.e., the nearer prototype) (exact binomial $p = .00016$, two-tailed). Clearly, Experiment 1's categorization training strongly reversed people's initial preferences.

Cluster dispersions are adjusted to maximize within-category activation and to minimize unwanted activation from items belonging to the opposing category. These dy-

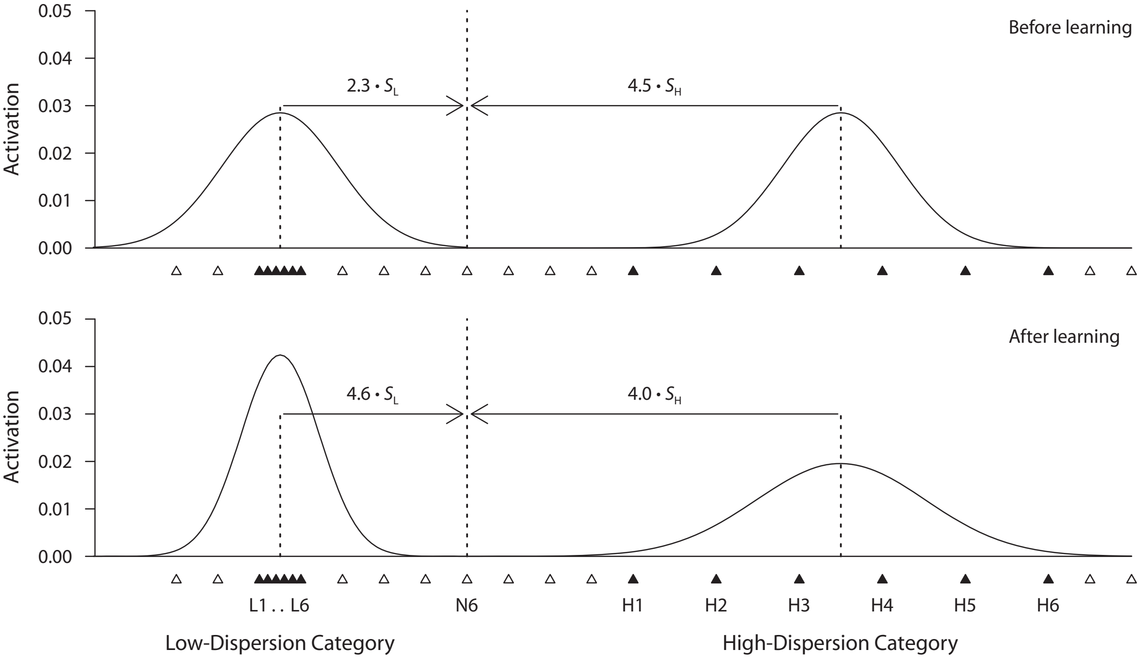

Figure 3. The activations of the clusters encoding the low- (Cluster L) and high-dispersion (Cluster H) categories in the mechanistic model are shown for each stimulus item in Experiment 1. The top panel shows that the variability for each cluster is equal before learning. An arrow indicates the number of standard deviations from the mean of each cluster to Border Item N6. Before learning, the border item is fewer standard deviations from Cluster L's center than from Cluster H's center, leading to greater activation and higher response probability for the low-dispersion category. The bottom panel shows that the opposite pattern arises after learning, due to the tightening of Cluster L and the widening of Cluster H (which make each cluster relatively more responsive to its category's members). After learning, the border item is more likely to be assigned to the high-dispersion category.

namics lead to learned standard deviations of 9.7 for the low-dispersion category and 22.7 for the high-dispersion category (averaged across simulations). Consequently, Item N6 more strongly activates the high-dispersion category's cluster after learning. These effects are illustrated in the bottom panel of Figure 3. The ratio of cluster activations for Stimulus N6 leads to a 70% probability of selecting the high-dispersion category, in close agreement with the empirical data. The operation of the mechanistic models mirrors that of rational models, with the one subtle difference being that the mechanistic model's estimates of mean and variance are made locally with regard to the current category representation.

EXPERIMENT 2

Both rational and mechanistic accounts captured Experiment 1's main finding: Key Border Item N6 was assigned to the high-variance category during transfer. Given the elegance, soundness, and nonarbitrary form of the rational accounts, one might question the value of a mechanistic account that requires consideration of unobservable learning processes and category representations. To the contrary, we argue that the worth of mechanistic accounts lies in these considerations and that consequent deviations from rationality (even ostensibly minor ones) can lead to important insights into human behavior.

In Experiment 2, we manipulated trial order to tease apart predictions for the mechanistic and rational models. Order effects have been extensively studied in category learning (e.g., Clapper, 2006; Medin & Bettger, 1994; Zaki & Homa, 1999). In fact, rational models have been developed to account for the effects of category drift (i.e., recency effects) for autocorrelated environments that change over time (Elliott & Anderson, 1995). Unlike these previous studies and modeling efforts, we consider how order can affect perceptions of variability along a single stimulus dimension. This is in contrast to the majority of ordering studies, which have focused either on recency effects or on detection of category patterns defined across multiple stimulus dimensions.

Experiment 2's design explored the key difference between the mechanistic and the rational models considered in Experiment 1. Unlike the rational models considered in Experiment 1, the mechanistic model updates its memory representation of each category in a local trial-by-trial fashion. The mechanistic model predicts that perceptions of category variability are based on trial-by-trial discrepancies between the current stimulus and the memory representation of the category (i.e., the position of the respective cluster). The rational models considered are not subject to the mechanistic model's processing limitations and, therefore, are not sensitive to this class of ordering effects.

In Experiment 2, members of one category appeared in an ordered fashion, so that successively presented members did not vary much from each other. In contrast, members of the other category were presented in a random fashion, as were members of both categories in Experiment 1. Globally, both categories in Experiment 2 had identical variability. However, the mechanistic model predicts that the discrepancy between the position of a category's cluster and the current stimulus will be smaller, on average, for the ordered category and, therefore, humans should treat the random category as more variable and assign Item N6 to it. This prediction is based on how cluster positions are updated in a local, trial-by-trial fashion. For the random category, the cluster position will fluctuate tightly around the true category mean, whereas for the ordered category, the cluster position will smoothly track the periodic oscillations created by the ordering manipulation (see Figure 5), leading to smaller average discrepancies and a lower estimate of category variability. To foreshadow Experiment 2's results, the predictions of the mechanistic model held.

Method

Forty-eight University of Texas undergraduates were tested. The procedure was the same as that in Experiment 1, except for the line lengths and order of stimulus presentation. Stimuli ranged from 60 to 180 pixels in one category and from 260 to 380 pixels in the other. Adjacent items differed by 5 pixels, resulting in 25 items per category. Whether the ordered category had longer or shorter lines than the random category was counterbalanced across subjects.

Every member of each category appeared exactly twice during training. The presentation order for this phase was determined by first generating a sequence for each category and then randomly interleaving these sequences in blocks of 10 (5 from each category sequence). The sequence for the ordered category was designed to reduce local variability. This sequence proceeded from the middle of the category distribution to the extreme (i.e., moving away from the category boundary), from the extreme to the category boundary (passing through the middle), and then from the boundary back to the middle of the category. More precisely, the sequence was generated by starting with the sequence O12, . . . , O1, O1, . . . , O25, O25, . . . , O13 and swapping each adjacent pair (excluding the first and last) with a probability of .5. Under this scheme, the items closest to Border Item N6 are presented after the items farthest away, so a simple explanation from recency effects works against our hypothesis. The presentation order for the random category was random, except for the first and last items, which were constrained to be R14 and R13, respectively (mirroring the ordered category). Figure 4 shows an example stimulus sequence.

The transfer stimuli consisted of lines of lengths (in pixels) 40 and 50 (novel items); 60, 90, 120, 150, and 180 (training items); 190, 200, 210, 220, 230, 240, and 250 (novel items); 260, 290, 320, 350 and 380 (training items); and 390 and 400 (novel items); with 220 as the critical Border Item N6. As in Experiment 1, the subjects completed two blocks of transfer, each with a random presentation order.

Results and Model Fits

The subjects were more likely to classify Border Item N6 into the random than into the ordered category. As is shown in Figure 5, averaged across the two transfer blocks, the subjects assigned the border item to the random category with a probability of .80, which is significantly greater than

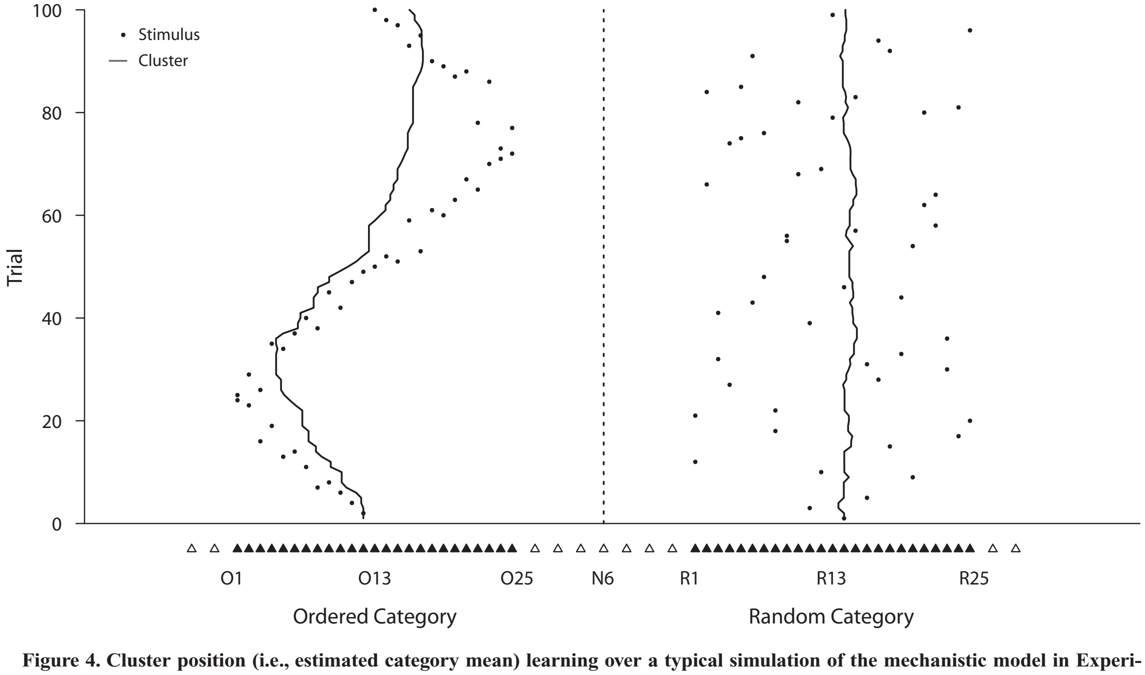

Figure 4. Cluster position (i.e., estimated category mean) learning over a typical simulation of the mechanistic model in Experiment 2. The horizontal axis denotes stimulus length, and the vertical axis captures the learning trial sequence. Each training stimulus is depicted by a solid triangle. The solid lines show evolving cluster positions. The cluster position for the ordered category follows the trajectory of the learning items. In comparison with the random category, this tracking leads to smaller differences between each stimulus and the current cluster position. Because of these smaller discrepancies, the model learns a lower variability for the ordered category and assigns Border Item N6 to the random category, in agreement with human subjects.

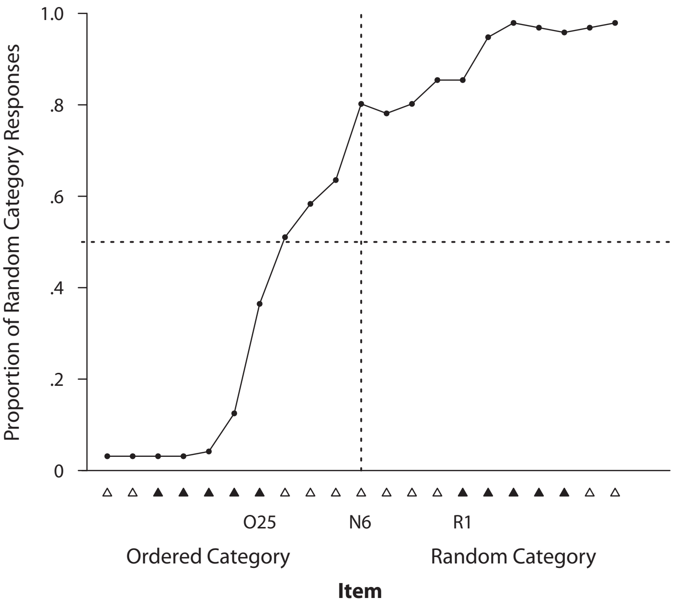

Figure 5. The probability of subjects' classifying each stimulus item as a member of the random category during the transfer phase of Experiment 2. Learning items are shown as dark triangles; novel items are shown as light triangles. Item N6 is midway between the nearest studied members (O25 and R1) of the ordered and random categories (and also midway between the prototypes of the two categories).

chance [$t(47) = 6.19, p < .001$]. In the first transfer block, more subjects (38 of 48) classified the border item into the random category than was expected by chance (exact binomial $p = .000062$, two-tailed). The same pattern (39 of 48) was found in the second transfer block (exact binomial $p = .000015$, two-tailed).

The mechanistic model was fit to Experiment 2's data, using the same parameter values as those used in Experiment 1's simulation. As was expected, the cluster mean for the ordered category tracked the stimuli, leading to lower average discrepancy between the cluster mean and each current stimulus, which in turn resulted in a lower variability for that cluster than for the cluster for the random category. This local effect, shown in Figure 4, resulted in average standard deviations of 17.9 for the ordered category and 25.6 for the random category after learning. Consequently, Item N6 more strongly activated the random category's cluster, leading to a 79% probability of selecting the random category, in close agreement with the human result.

DISCUSSION

In Experiment 1, people appeared sensitive to category variability and assigned a transfer item lying between two categories to the higher variability category. This finding suggests natural accounts from both mechanistic and rational perspectives. These accounts largely converge, in that both assume that people learn the mean and variability of each category and use that information to classify new items. One distinguishing and nonrational aspect of the mechanistic account is that estimates of category mean and variance are made in a trial-by-trial fashion. Instead of calculating an unbiased estimate of these quantities, the mechanistic model employs local learning rules that are driven by discrepancies between the memory representation of the category (i.e., the cluster) and the current stimulus.

This departure from rationality might seem modest, but it was the basis for a surprising prediction that was confirmed in Experiment 2. In Experiment 2, both categories had equal variance, but one category was ordered semiregularly, so that differences between the stimuli on successive trials were small. The mechanistic model predicted that this ordering would create an illusion of low variability for the ordered category, since the discrepancy between each presented stimulus and the current category representation was relatively small. Accordingly, human subjects assigned the border item at transfer to the randomly ordered category. Overall, these empirical and modeling results suggest that people estimate variability by making incremental adjustments to memory representations on the basis of local comparisons. These results also suggest that consideration of mechanistic models, with their accompanying processes and representations, is a fruitful research strategy, particularly when departures from rationality are considered.

One persistent criticism of the mechanistic approach is that multiple mechanisms can give rise to the same behavior (Townsend, 1974). Proponents of the rational approach argue that grounding models in the structure of the environment provides additional constraints. Although incorporating additional constraints is desirable, we find the claims that there are privileged and unambiguous facts about the environment to be dubious. Any environment of sufficient complexity can be characterized in a number of different ways. Assumptions about what information people monitor, the dynamics of the environment, and associated rewards can vary. When a rational analysis fails, these assumptions are altered until the desired result is achieved (as in Step 6 of Anderson's, 1990, rationality framework).

Importantly, these attacks on mechanistic accounts ignore the substantial constraints that such a perspective provides. Mechanistic accounts are not made in a theoretical vacuum but are informed by existing models and behavioral findings. Current thinking on the nature of our cognitive architecture (e.g., capacity-limited working memory) provides grounding, and the successes and failures of related models provide lessons for future models. This was the case in formalizing the mechanistic model presented here, in light of the substantial evidence for similarity-based representations and error-driven learning, coupled with the specific failures of prototype and exemplar models in explaining Experiment 1. In practice, the mechanistic approach may offer more constraints than the rational approach's *first principles* orientation, which emphasizes de novo analysis of the current task and environment for each application.

We are not suggesting that there is not a rational account of Experiment 2's results. There are likely an infinite number of possible rational explanations. For example, a rational account that assumes categories in the environment steadily drift could be made consistent with Experiment 2's results. Along these lines, Elliott and Anderson (1995) presented a rational model that makes these assumptions in regard to estimating a category's mean. Interestingly, their model's disproportionate weighting of recent items leads it to assign the border item to the ordered category in Experiment 2, whereas human subjects tended to assign the border item to the random category.

The key metascientific question is whether successful rational explanations would come to the forefront prior to specifying Experiment 2's design. Following Experiment 1, we specified the most straightforward and readily suggested rational and mechanistic accounts. Focusing on the rational explanation of Experiment 1 would not have led to Experiment 2, whereas considering *how* the mechanistic model accounted for Experiment 1's results did motivate Experiment 2.

One possibility is that rational explanations, although illuminating and satisfying, largely serve as just-so stories that are constructed after interesting behavioral findings present themselves. According to this view, rational analyses are more likely to follow from mechanistic explanations than vice versa. Perhaps one example of this progression is from the RULEX (Nosofsky, Palmeri, & McKinley, 1994) model of hypothesis generation and testing to Boolean complexity (Feldman, 2000). RULEX specifies how people search for Boolean rules by beginning with simple rules and progressing toward more complex rules when simple rules fail. Boolean complexity preserves many of these insights, albeit in a more abstract form that does away with RULEX's proposed search and memory processes. Instead, Boolean complexity offers a well-formulated metric that is derived through a rational analysis.

Although our discussion has been provocative and heavily tilted in favor of mechanistic approaches, we do not wish to suggest that rational analysis does not have its place. Here, we suggest that mechanistic models can guide rational analyses. Likewise, rational analyses can guide the development of mechanistic models. A rational analysis can uncover the principles that mechanistic models approximate and bring into focus how a mechanistic model deviates from rationality. Experiment 2's design was motivated by such considerations. In addition, consideration of what environmental assumptions would rationally justify behaviors exhibited by mechanistic models can provide insight into our cognitive environment, such as the idea that real categories drift over time, as suggested by a post hoc rational analysis of Experiment 2. Researchers in the field are likely to make progress when intellectual effort is devoted to both approaches. Given the recent tilt toward rational approaches, we would like to end by encouraging the researchers in the field not to shy away from mechanistic explanations. If the main question we are trying to answer is how the mind works, we should not fear directly addressing this question by developing and evaluating mechanistic models.

AUTHOR NOTE

This work was supported by Grants AFOSR FA9550-04-1-0226 and NSF CAREER 0349101 to B.C.L. and by Grant NIMH NRSA F32-MH068965 to M.J. Correspondence concerning this research should be addressed to B. C. Love, Department of Psychology, University of Texas, 1 University Station A8000, Austin, TX 78712-0187 (e-mail: brad_love@mail.utexas.edu).

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

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

Anderson, J. R. (1991b). Is human cognition adaptive? *Behavioral & Brain Sciences*, **14**, 471-484.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, **3**, 57-65.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Where next? *Trends in Cognitive Sciences*, **10**, 292-293.

Clapper, J. P. (2006). When more is less: Negative exposure effects in unsupervised learning. *Memory & Cognition*, **34**, 890-902.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, **29**, 1165-1175.

Elliott, S. W., & Anderson, J. R. (1995). Effect of memory decay on predictions from changing categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 815-836.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, **407**, 630-633.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 234-257.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 59-100). Cambridge: Cambridge University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, **17**, 767-773.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, **33**, 289-302.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Love, B. C., & Jones, M. (2006). The emergence of multiple learning systems. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 507-512). Mahwah, NJ: Erlbaum.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, **111**, 309-332.

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 301-321.

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

Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, **1**, 250-254.

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Ono, H. (1967). Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. *Perception & Psychophysics*, **2**, 201-207.

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

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.

Rodrigues, P. M., & Murre, J. M. J. (2007). Rules-plus-exception tasks: A problem for exemplar models? *Psychonomic Bulletin & Review*, **14**, 640-646.

Sakamoto, Y., Matsuka, T., & Love, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th International Conference of Cognitive Modeling* (pp. 261-266). Mahwah, NJ: Erlbaum.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral & Brain Sciences*, **24**, 629-640.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133-186). Hillsdale, NJ: Erlbaum.

Zaki, S. R., & Homa, D. (1999). Concepts and transformational knowledge. *Cognitive Psychology*, **39**, 69-115.

NOTE

1. All of these models assume that category members are distributed according to the Gaussian (i.e., normal) distribution. This common choice is motivated by a variety of considerations, ranging from the nature of noise in the nervous system to the general structure of categories in our environment. Rational models can be formulated using other distributions when such distributions better conform to the structure of a particular domain (cf. Griffiths & Tenenbaum, 2006).

Interestingly, classic prototype models and exemplar models have difficulty accounting for Experiment 1's results. Prototype models represent each category by its prototypical (or average) member (Reed, 1972). Exemplar models represent categories by storing all encountered examples (Medin & Schaffer, 1978). Both models classify new instances on the basis of their relative similarity to these stored category representations. Both standard prototype and exemplar models strongly predict that subjects will classify Border Item N6 into the low-dispersion category, because the same similarity metric is used for the low- and the high-dispersion categories and the prototype for the low-dispersion category is closer to N6, as are the exemplars forming the low-dispersion category.

(Manuscript received June 29, 2006;
revision accepted for publication April 11, 2008.)