Journal of Experimental Psychology: General
1999, Vol. 128, No. 2, 107-130

0096-3445/99/$3.00

Task Representations, Strategy Variability, and Base-Rate Neglect

Marsha C. Lovett
Carnegie Mellon University

Christian D. Schunn
George Mason University

The authors present a new model called RCCL (pronounced "ReCyCLe"; Represent the task, Construct a set of action strategies consistent with the task representation, Choose from among those strategies according to their success rates, and Learn new success rates for the strategies based on experience). The model explains the different ways in which people combine base-rate and case-specific cues to produce choice. It also makes additional predictions regarding variability in people's choices over time. Experiment 1 tested 58 college-age students in a problem-solving task and showed that task representations can be influenced by feedback from the environment, producing changes in base rate and cue sensitivity. Experiment 2 tested 80 college-age students in a delayed match-to-sample task and showed that variations in the format of a task can lead to different representations, which in turn produce much different base-rate and cue sensitivities. Moreover, both experiments showed systematic variability in choice over time in ways predicted by the model.

The difficulty of making choices in an uncertain world is that it is impossible to know in advance which choice will lead to success. For example, in driving to the airport, any particular route may be congested with traffic or blocked by an accident. In such situations, people must base their choices on the information they have available about the likely success of each alternative. Such information can be cast in terms of two types: (a) the overall base rate of success of each alternative (e.g., how often each route is not congested) and (b) information specific to the current case that is predictive of a certain alternative's success (e.g., the time of day indicating that there will likely be congestion on a particular route). Even with both types of information, choice is not simple: Each type of information is often only partially predictive of whether an alternative will succeed, and the two types of information may indicate opposite choices. With both kinds of information available, how do people make their choices? That is, how do people combine overall base-rate information with case-specific information?

This issue has been examined using text-based problems that explicitly present base-rate and case-specific information separately and require participants to predict the probability that one alternative will succeed. Studies using this paradigm have typically showed *base-rate neglect* (e.g., Ginossar & Trope, 1987; Lyon & Slovic, 1976; Kahneman & Tversky, 1973). Base-rate neglect is said to occur when people's predictions do not adequately take into account the overall probabilities (or base rates) of the various alternatives. For example, Tversky and Kahneman (1982) found that when given the information that (a) 85% of the taxicabs in a city are green and (b) a witness with 80% reliability saw a blue cab, participants typically estimated that the cab seen by the witness was blue with an 80% probability (although the correct answer is 41%). The participants underestimated the impact of the base rate (85% green) in favor of the case-specific information (80% reliability for seeing blue).

Recently, however, the severity and robustness of base-rate neglect have been called into question (see Koehler, 1996, for a review). For instance, base-rate neglect is reduced or even eliminated when the problem's wording is manipulated to emphasize certain features such as the independence of the two kinds of information (Macchi, 1995), the perceived relevance of the base rates (Ajzen, 1977; Bar-Hillel, 1980; Beckett & Park, 1995; Birnbaum & Mellers, 1983; Carroll & Siegler, 1977; Fischhoff, Slovic, & Lichtenstein, 1979), or the random sampling of cases (Gigerenzer, Hell, & Blank, 1988). In addition, participants do not exhibit base-rate neglect when they are given a frequency-based problem statement rather than a probability-based statement (Gigerenzer & Hoffrage, 1995). Thus, base-rate neglect is not a universal phenomenon but one that is specific to certain presentation formats.

These findings of base-rate neglect and base-rate sensitivity are all based on situations in which information is presented explicitly via text. In many situations, however, people must gather information about the base rates of success and the predictiveness of case-specific cues through personal experience. For example, in learning to navigate in a new city, a driver learns over years of experience which routes are most successful and which cues are indicative of

Marsha C. Lovett, Department of Psychology, Carnegie Mellon University; Christian D. Schunn, Department of Psychology, George Mason University.

This research was supported by Grant N00014-95-1-0223 from the Office of Naval Research. We thank Robert Siegler, Martha Alibali, Steve Ritter, Doug Thompson, Bethany Rittle-Johnson, and Zhe Chen for several discussions that contributed to this work; Adam Goodie for his comments on a previous version of this article; and Jessica Phillips-Silver for conducting the experiments and coding data.

Correspondence concerning this article should be addressed to Marsha C. Lovett, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to lovett@cmu.edu.

success for each alternative. This raises the question of how people combine base-rate and cue-specific information when they must induce it over the course of many trials.

To examine choice in such situations, researchers have used a paradigm in which participants *experience* both types of information in a trial-by-trial sequence of cases. In particular, case-specific information is provided for each trial, after which participants make a choice and receive feedback. This feedback, accumulated across trials, presents the base rates (i.e., the proportion of trials for which each choice is correct) and the predictiveness of cues (i.e., the proportion of trials for which a given choice is correct in the presence of a given cue). In this paradigm, people learn the task contingencies by experience and express their predictions through choice.

Results show that in this experiential situation, people exhibit striking sensitivity to base rates (Estes, Campbell, Hatsopoulos, & Hurwitz, 1989; Gigerenzer et al., 1988; Gluck & Bower, 1988; Maddox, 1995; Manis, Dovalina, Avis, & Cardoze, 1980). For example, Estes et al. (1989) and Gluck and Bower (1988) used a medical diagnosis task to study experiential base-rate phenomena. In this task, participants were asked on each trial to read a list of symptoms exhibited by a hypothetical patient and then to classify that patient's disease based on the symptoms. The overall base rates of the two diseases were set at .25 and .75, making one disease "rare" and the other "common." After approximately 200 trials, participants' disease classifications for the various training configurations revealed base-rate sensitivity; participants preferred the common disease in their classifications. In particular, participants' classification proportions for the different symptom configurations conformed almost exactly to a normative standard. These results suggest that people show base-rate-sensitive behavior when they learn and express their predictions in a nonverbal way. Explanations of such results rest on the notion that learning base rates implicitly (via experience) and making choices implicitly (via behavior) tend to outperform explicit learning and explicit computation (Spellman, 1996).

Given these results, researchers have been tempted to conclude that experience is the key to base-rate sensitivity, that is, people are sensitive to base rates when the problem is presented experientially but not necessarily when the problem is presented via text. There is more to this story, however. Several studies contradict this generalization by showing that base-rate neglect can occur in direct-experience situations (e.g., Gluck & Bower, 1988; Goodie & Fantino, 1995, 1996; Medin & Edelson, 1988). Some of the cases of base-rate neglect occur only for test trials that are distinct from training (Estes et al., 1989; Gluck & Bower, 1988; Medin & Edelson, 1988), but one particular task has been shown to produce base-rate neglect throughout repeated training problems (Goodie & Fantino, 1995, 1996). Thus, the entire continuum of base-rate effects exists within the experiential paradigm just as it does for the text-based paradigm. However, unlike the text-based situations, researchers do not know what factors influence people's sensitivity to base-rate information in direct-experience situations. Moreover, they do not know how people combine base-rate and cue-specific information in such direct-experience situations.

In this article, we argue that an individual's *task representation* modulates the capacity to learn base-rate and cue-predictiveness information. The term *representation* has been used previously in the literature to explain base-rate sensitivity in verbal formats (e.g., Gigerenzer & Hoffrage, 1995; Koehler, 1996). We define a task representation as the set of stimulus features an individual uses to encode the task environment. For example, when choosing a route to take to work, a driver may encode the time of day and whether each option involves highway driving. This task representation, however, includes only a subset of all possible task features. It therefore acts as a filter on what can be learned from experience. For instance, in the previous example, the driver did not include weather conditions in the task representation and thus could not learn about the predictiveness of this cue. On the other hand, the driver did represent the highway status of the different options and thus could learn differential base rates of success for highway versus nonhighway driving.

Note that our use of the term *representation* differs from other uses in the literature. For instance, Gigerenzer and Hoffrage (1995) used "the general term *information representation* and the specific terms *information format* and *information menu* to refer to *external* representations, recorded on paper or on some other physical medium" (p. 685). For example, they manipulated whether textual problem descriptions would present events in terms of either probabilities or frequencies and then examined the impact on people's base-rate sensitivity. In contrast, we use the term *task representation* to refer to *internal* representations, specifically the set of features an individual uses to encode the stimuli of the task. Both internal and external representations act as general filters on the information an individual obtains from a task. In textual paradigms, they influence the kinds of information transformations that people will apply. In experiential paradigms, they influence which task contingencies people can learn and hence the choice preferences they will exhibit. The external representation determines which information is easily available, and the internal representation selects a subset of that information for processing. For example, an individual may not internally represent all the information presented externally and may also change over time which features are internally represented. It is the internal task representation that ultimately determines behavior and thus is the focus of this article.

The RCCL Model

We propose a process model called *RCCL* (pronounced "ReCyCLe") that specifies how task representations can influence choice in experiential base-rate situations (see Figure 1). The four main stages of processing in RCCL are as follows: (a) Represent the task, (b) Construct a set of action strategies consistent with that task representation, (c) Choose from among those strategies according to their success rates, and (d) Learn new success rates for the

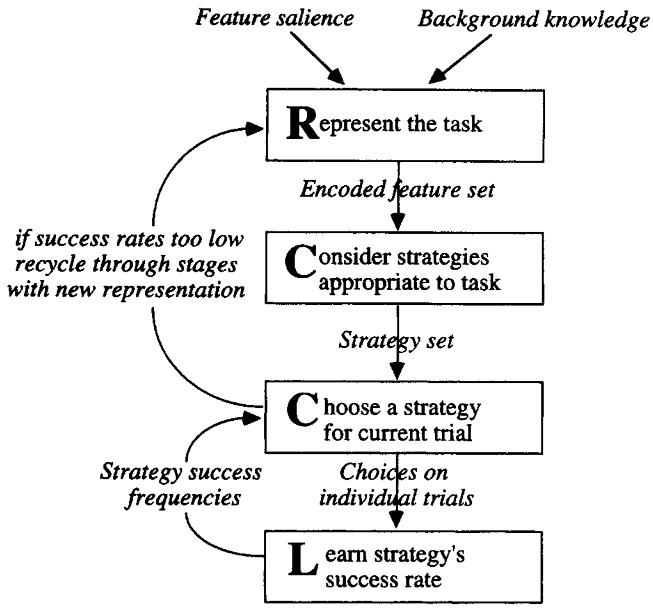

Figure 1. The stages of processing in the model.

strategies based on experience. The primary theme underlying the RCCL model is that a task representation constrains the set of strategies an individual will use for taking actions in the task environment. Making choices according to the learned success rates of a certain set of strategies enables RCCL to produce base-rate sensitivity or base-rate neglect in direct-experience situations; sensitivity arises only when the constructed strategies include stimulus features that are important to success in the task. The RCCL model also includes recycling through the aforementioned processes when the current representation and strategies lead to low success rates. This implies that an individual's task representation need not be static but that it can develop with experience. The following paragraphs describe the RCCL model's processes in further detail.

The first stage in RCCL, represent the task, posits that an individual's background knowledge and the salience of different features in a task are used to select a certain *subset* of features to be included in the task representation. These features provide a means of filtering information so that it can be attended to and organized. In particular, trials that have the same values on represented features can be grouped together, aggregating over any varying, nonrepresented features. For instance, because the driver in the earlier example represented time of day and highway driving, he or she would consider all experiences of morning highway driving together (even if they differed in terms of weather) but would distinguish between morning highway driving and midday highway driving (because they differ in terms of the represented feature, time of day).

The features of a task representation filter and organize incoming information, but the resulting partition of experiences is latent; it is only a knowledge base and itself does not provide a means for taking action. In the second stage of RCCL, construct strategies, the features of the task representation are combined in various ways to generate a set of strategies for action. Again, prior knowledge provides a constraint in that each individual will generate only a subset of the possible combinations. We take each strategy to be expressed in the form "if ⟨condition⟩ then ⟨action⟩." Represented features that describe cues in the task environment fill the condition slot, and represented features that describe the various choices fill the action slot. For instance, given the particular representation of driving exemplified previously, the following strategy could be constructed: "If the time of day is morning, then take the highway route." This strategy uses the represented feature "time of day" in the condition slot and the represented feature "highway status" in the action slot. In general, a constructed strategy specifies a particular action to be taken in the presence of a particular cue, enabling cue-sensitive choice.

When multiple strategies' conditions are met by the current situation, the individual must choose a single strategy to determine what action to take. The third stage of RCCL, choose among strategies, handles this situation by choosing the strategy with the highest estimated rate of success. Each strategy's estimated success rate is learned via experience (see Stage 4). Given a set of competing strategies, Stage 3 assumes some noise in the estimation and choice process such that the most successful strategy is often but not always used. Note that this noisy selection process can lead to probability matching, a specific kind of base-rate sensitivity in which people choose each option a proportion of the time equal to its probability of success (Lovett & Anderson, 1995).

Stage 4 of RCCL, learn the strategies' success rates, specifies how the estimated success rates, used in Stage 3, are learned. The RCCL model records success information for each strategy so that it can adaptively choose between competing strategies. At first, each strategy's estimated success rate takes a value based on the solver's prior knowledge. Then, with experience, this estimate moves toward the experienced success rate. This learning mechanism leads the RCCL model to produce gradual changes in its success estimates as it gains experience in the task. These changes in turn lead to changes in choice preference: The model will learn how often each strategy is successful and then choose a given strategy in proportion to its relative success rate. Thus, the model posits sensitivity to *strategy-specific base rates.* Sensitivity to *overall base rates* of the task, however, should occur only when the individual's task representation and strategies include the relevant features of the task. For example, suppose that in our driving task the overall base rate of success for right-turn routes is greater than that for left-turn routes. According to our model, the driver who does not represent this left–right feature will not be able to learn these different base rates and will instead make choices that are sensitive to the base rates of features included in his or her task representation. This contrasts with an implicit learning view that claims people can learn (and behave in accord with) base rates without awareness of the predictive features (e.g., Berry & Broadbent, 1984). Our model posits that task representations play a pivotal role in explaining the source of base-rate effects: Only base rates for the features explicitly included in an individual's representation will be learned and will govern behavior. Although the relative success rates of different strategies can be learned implicitly in our model (e.g., an individual may not be able to accurately express explicitly the probabilities for success), all the features included in the task representation will be explicitly accessible during problem solving.

Task representations also play another important role in the RCCL model: They predict changes in feature use over time under certain conditions. When the strategies generated under the current task representation are all performing poorly, then the RCCL model initiates a change in task representation, by adding features, deleting features, or both. New strategies (using these new features) are then created, and their relative success base rates are learned through experience. Thus, as RCCL recycles through the four stages with new task representations, new base-rate sensitivity or neglect can result. This second layer of adaptivity suggests that over time, an individual will move toward the task representation and strategy set that produce the highest success rates.

The mechanisms specified by the RCCL model lead to several predictions about the integration of cue and base-rate information to produce choice behavior:

1. *Salient features of the task will influence initial cue use.* This prediction stems from the first and second stages of RCCL. Salient features of the task are included in the task representation and then combined, on the basis of prior knowledge, to form strategies. Because these strategies are the basis for taking action in the task, people's choices will be sensitive only to the features included in their strategies. This contrasts with an implicit memory account that can predict complete learning even in the absence of explicit encoding of relevant features.

2. *Under a given task representation, people will learn to prefer the strategies that have higher base rates of success.* This prediction stems from the third and fourth stages of the RCCL model that claim people choose from among competing strategies based on their learned rates of success. A strategy's initial estimate of success gets adjusted according to how often it actually leads to success when used. The updated estimate is then used to determine how often that strategy will be chosen over its competitors. This process allows the RCCL model to demonstrate base-rate sensitivity: An action will be taken more often the more successful its corresponding strategy. Thus, contrary to some previous findings of apparent base-rate neglect, RCCL clearly predicts that one should always find base-rate sensitivity, as long as the relevant features are included in the individual's task representation.

3. *People will drop cues that prove to be irrelevant.* This prediction stems from two aspects of the model: First, strategies that use relevant features will be more successful than strategies that use irrelevant features, and over time the less successful strategies will no longer be used. Second, task representations that primarily include irrelevant features will lead to low overall success rates, which will in turn produce task representation change.

4. *More representation and strategy change will occur in tasks with low success rates.* This prediction stems from the process of representational change included in the RCCL model. When the strategies generated from the current representation all lead to low success rates, the task is re-represented. This change process may involve adding new features to the task representation and deleting old ones. The four stages of the RCCL model are cycled through again, so a new set of strategies (with newly learned success rates) will determine choice behavior.

Although these predictions may seem intuitive and unsurprising, note that we know of no other current model that makes such a combination of predictions for choice behavior in problem-solving tasks. Some models of categorization make predictions about feature salience and use, similar to Prediction 1 (e.g., Anderson, 1991; Goldstone, Medin, & Halberstadt, 1997; Kruschke, 1996; Nosofsky, 1984), and some other models make predictions about how choices are based on success, similar to Prediction 2 (e.g., Anderson, 1993; Busemeyer & Myung, 1992; Estes et al., 1989). The RCCL model, however, includes predictions on both of these issues within a single framework. Given that people facing a new problem-solving task must both generate new strategies for choosing actions and learn which of those strategies are preferable, RCCL's integration of these two processes is a critical step toward understanding problem-solving choice.

Although Predictions 1 and 2 may be separately predicted by other models, Predictions 3 and 4 are unique to the RCCL model. Previous models of categorization and problem solving have tended to assume a fixed representation over time, whereas the RCCL model predicts that representation change should occur regularly and that this is a fundamental component of learning in a domain. This contrasts sharply with an implicit learning model that assumes that explicit task representation change is unnecessary for learning the desired cue–action pairings. The RCCL model not only predicts that representation change will occur and is important for learning but it also makes a particular prediction for when representation change should occur. Although the RCCL model is unique in making any prediction for this issue, other possible outcomes are also possible and plausible. For example, it may be that features are added and deleted stochastically to and from the task representation uniformly over time (e.g., with a genetic algorithm) rather than primarily when success rates are low.

In Experiment 1 we tested all four predictions. In Experiment 2 we tested additional predictions of the RCCL model, focusing on the role of task representations in various base-rate effects.

Experiment 1

To test the four predictions of the RCCL model, we used a problem-solving task in which the first step in each problem solution involved making a choice. We focused our study on how base-rate and cue-specific information would influence this choice. The task met the requirements of an experiential base-rate task in that case-specific cues and feedback were presented on each trial. Moreover, it was a useful experimental task because participants' choices were observable and easily identifiable and because both base rates and cue predictiveness could be experimentally manipulated. Additionally, because of the task's problem-solving nature, the way participants encoded each trial and how they made their choices were a natural part of the solution process. That is, each trial was not explicitly presented as a set of features and a set of options; instead, the trial's features and options were embedded in the context of a given problem to be solved. This could lead participants to approach the task in a way that was more similar to the way they approached real-world choice situations. However, the task was simple enough that experimental control of the task would be high and operationalizing a measure of choice was straightforward.

The problem-solving task we chose is called the Building Sticks Task (BST; Lovett & Anderson, 1996). For a given BST problem, solvers must add and subtract an unlimited supply of three different-sized building sticks to create a stick of the desired length (see the top of Figure 2). BST problems can be solved by one of two procedures. The *undershoot procedure* involves starting with a building stick that is shorter than the desired stick and then lengthening that stick by additional building stick lengths until the desired stick's length is reached. In contrast, the *overshoot procedure* involves starting with the building stick that is longer than the desired stick and then shortening that stick by the other building stick lengths. As Figure 2 indicates, participants implicitly choose between these two solution procedures in their first step.

For example, suppose that the desired stick was 14 units long and that Sticks A, B, and C were 4, 17, and 6 units long, respectively. To obtain the desired stick length of 14 units, the participants might start with Stick B (17 units) and

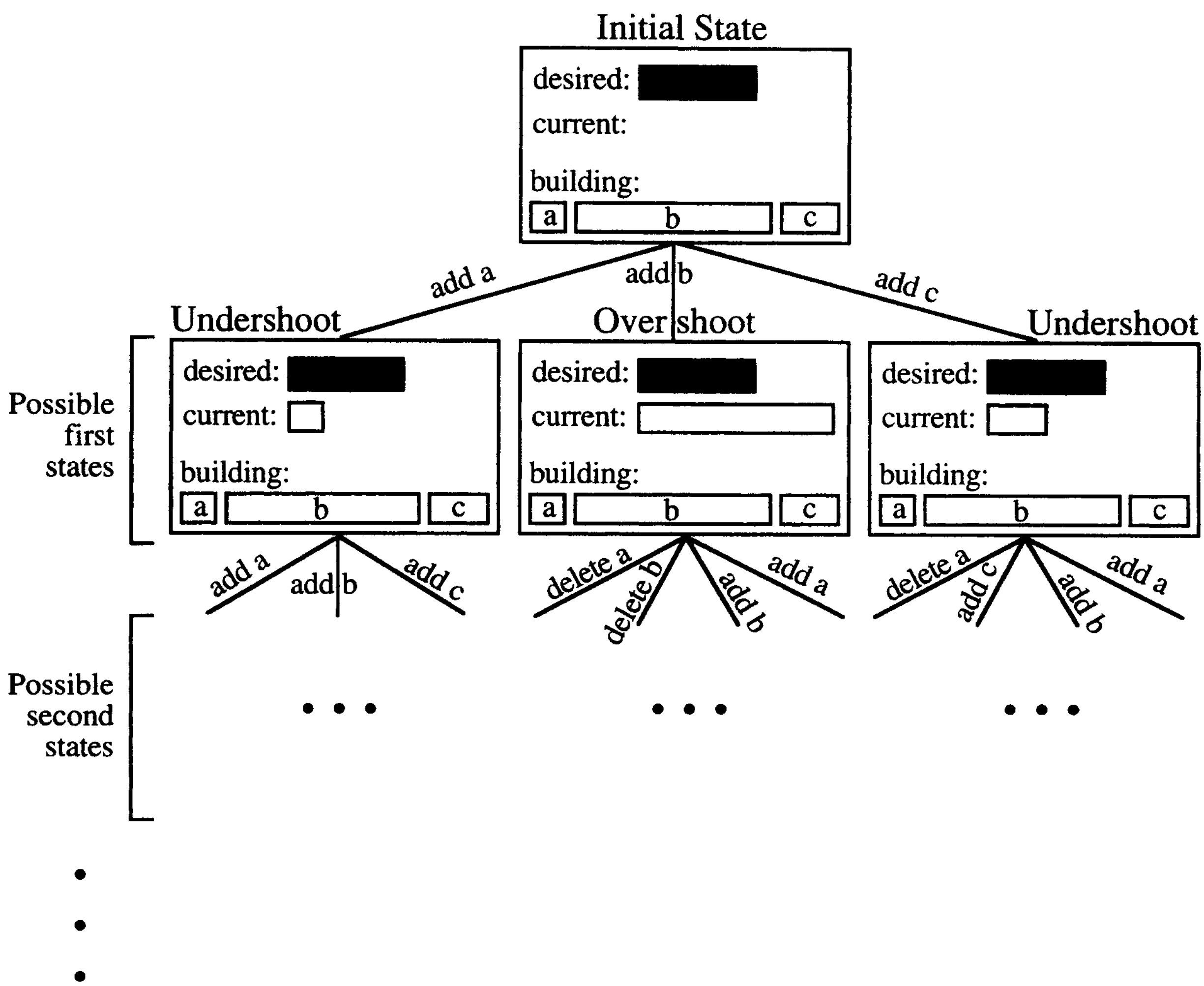

Figure 2. Initial and subsequent states in the Building Sticks Task (BST). Each BST problem includes three building sticks and a desired stick. The solvers' task is to build a current stick (initially Length 0) that matches the desired stick in length by adding and subtracting various combinations of the building stick lengths. A solver's solution strategy can be categorized into undershoot or overshoot strategies according to the first move.

remove segments (the overshoot procedure), or the participants might start with Stick C (6 units) and add more segments (the undershoot procedure). In this example, a solution can be obtained only by the undershoot procedure ($C + C + A = 6 + 4 + 4 = 14$). The overshoot procedure will not work because subtracting Lengths A and C from B will never lead exactly to a stick of 14 units long. Of course, in other problems the overshoot procedure may be the correct one to use.

Note that the participants were never given the exact numerical lengths of the sticks; the example just given was used for expository purposes. In the experiment, participants had to visually estimate the length of each stick. This prevented them from being able to solve the task algebraically and forced them to try a strategy (i.e., to make their choices externally) to determine whether it would work.

It was possible to vary base rates and cue predictiveness independently in this task. First, we discuss how these two types of information can be varied individually; we then show that such manipulations can be done independently. The manipulation of base rates is straightforward: We design each problem to be solvable by either undershoot or overshoot (but not both) and then vary the proportion of problems with each solution. To vary cue predictiveness, we design problems with certain feature patterns and then vary the proportion of problems with a given pattern that are solvable by one procedure. If all or almost all of the problems with a certain feature pattern are solved by one procedure, then that feature pattern is predictive of that procedure's success. On the other hand, if half of the problems with a certain feature pattern are solved by one procedure and the other half of those problems are solved by the other procedure, then that feature pattern is not predictive (i.e., seeing that feature pattern offers no information to help decide which procedure will solve the problem).

Figure 3 shows the feature pattern that we manipulated to be predictive or not predictive across conditions in this experiment. We call it the *relative length cue.* For the problems in the top row, Building Stick C is closer in length to the desired stick than is Building Stick B. This pattern of lengths provides a cue to use the undershoot procedure (i.e., to choose Building Stick C in the first step) because doing so gets closest to the problem goal. The problems in the bottom row of Figure 3 have the opposite length pattern: Building Stick B is closer in length to the desired stick than is Building Stick C. This provides a cue to use the overshoot procedure (i.e., to choose Building Stick B in the first step) because doing so gets closest to the problem goal. The RCCL model predicts that this cue will influence participants' initial choice because it makes use of salient features in the task (i.e., the lengths of the desired stick, Building Stick B, and Building Stick C) and combines them in a meaningful way based on prior knowledge (e.g., a hill-climbing heuristic that aims to get as close to the goal as possible with each step).

Figure 3 also shows how problems with the same relative

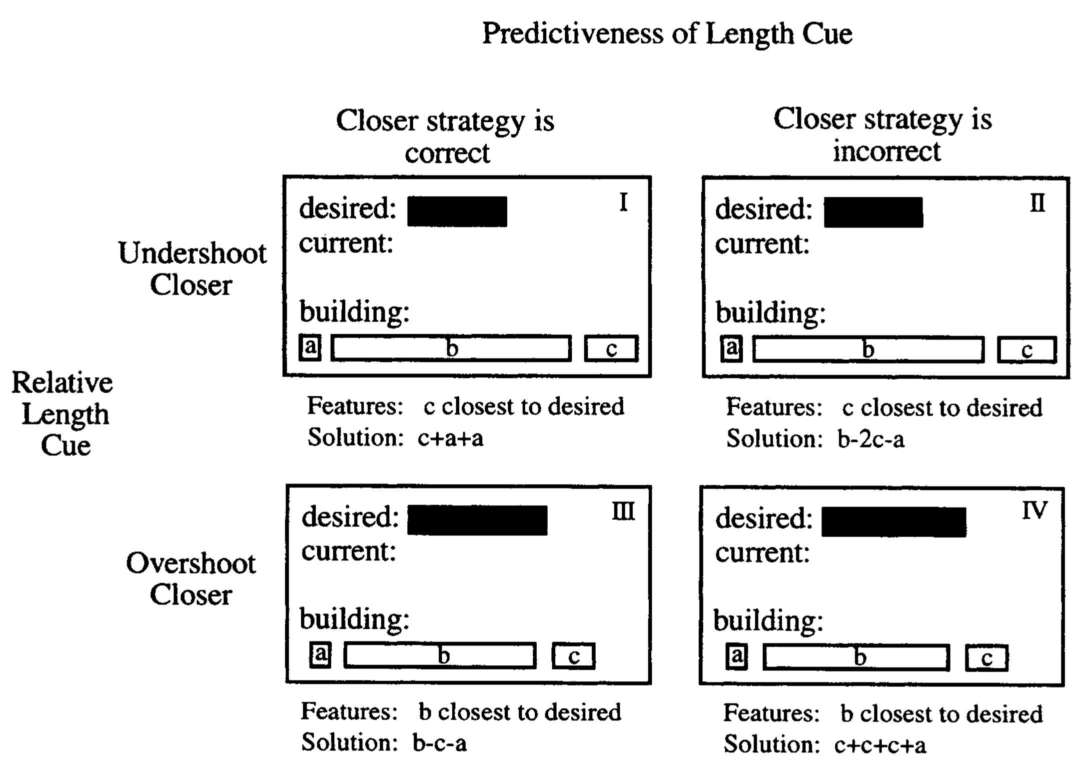

Figure 3. The four problem types (undershoot closer and overshoot closer × closer strategy is correct and closer strategy is incorrect) used as training problems in Experiment 1. These examples highlight that the length cue was not always predictive of the correct procedure.

length cue can be paired with either an undershoot or overshoot solution. Specifically, the two problems in each row have almost identical feature patterns but are solved by different procedures. By using these four problem types in different proportions, we can design problem sets that independently vary the base rates of undershoot versus overshoot success and the predictiveness of the relative length cue.

In Experiment 1 we manipulated the relative base rates of overshoot versus undershoot and the predictiveness of the relative length cue to test the four predictions of the RCCL model. First, initial cue use in the BST task should be determined by feature saliency. Specifically, participants should be initially sensitive to the lengths of the choices relative to the desired stick length. Second, participants should learn the relative rates of success of the overshoot and undershoot procedures because we expected that their task representations would distinguish between these procedures. Third, participants should stop using the relative length cue if it proves not to be predictive. Fourth, when the strategies generated under the initial task representation are relatively unsuccessful, participants should change their task representations and strategies (i.e., stop using the relative length cue).

Method

Participants. Fifty-eight Carnegie Mellon University undergraduates participated for course credit and were randomly assigned to one of four conditions.

Design. Four experimental conditions (2 × 2 design) differed in the set of training problems that participants solved. First, we manipulated the overall base rates of success of the two solution procedures. In the "biased base-rates case," 70% of the problems were solved by one procedure and 30% of problems by the other (with the procedure assigned to be more successful counterbalanced across participants). In the "not-biased case," 50% of the problems were solved by overshoot and 50% by undershoot. Manipulated orthogonally to these base rates was whether the relative length cue would be predictive of a certain procedure being successful. The cue was considered "predictive" when 80% of the undershoot-closer problems were solved by undershoot and 80% of the overshoot-closer problems were solved by overshoot. It was considered "not predictive" when 50% of the problems of each cue type were solved in accordance with their cue. The four conditions resulting from this factorial combination are labeled biased/predictive, biased/not predictive, not biased/predictive, and not biased/not predictive, in which the first term describes the base rate of the condition and the second term describes cue predictiveness.

Stimuli. All training problems were designed to have extreme relative length cues. That is, the relative stick lengths would be set so that Stick C was much closer to the goal than Stick B or vice versa. In addition, we designed each problem to have a "twin" version that looked highly similar in terms of its stick lengths (e.g., within a few pixels) but that was solved by the alternate procedure. (The same-row entries in Figure 3 are examples of such "twins.") This made it easy to manipulate whether a problem with a certain length cue was solved in accordance with its cue; we included only one or the other twin.

To create the problem set for each condition, we included a certain proportion of each of the four problem types shown in Figure 3. We constrained each problem set to include half undershoot-closer problems and half overshoot-closer problems. In this way, every participant would see the same distribution of problem features. For a condition in which the relative length cue was to be predictive, the proportion of problems whose solution matched the cue (Type I + Type III from Figure 3) was set to .80; for the not-predictive case, this proportion was set to .50. For the biased base-rate manipulation, the proportion of problems solved by the more successful procedure (e.g., Type I + Type IV from Figure 3) was set to .70, and, for the not-biased condition, it was set to .50. Table 1 shows the resulting proportions we used.

Procedure. At the beginning of the experiment, a computer tutorial provided participants with instructions in the BST and examples of problems solved by the undershoot and overshoot procedures, although the two strategies were not given explicit labels. The experiment itself consisted of 80 training problems. Participants worked on each of these problems until a solution was reached (i.e., until the length of the stick they were building matched the length of the desired stick). Participants were encouraged to use a "reset" button when they wanted to erase the current stick they were building and start a new solution attempt.

In addition to the 80 training problems, participants were given a set of 10 test problems immediately before and after the training set. The purpose of these test problems was to measure initial and final strategy use (i.e., cue use and overall rates of use of each strategy) without providing additional feedback about the success rates of the strategies during the test. We expected that participants would change their strategy use if feedback were provided. On these test problems, participants were asked to take only the first step on the way to a solution; after they did so on a given problem, the screen was automatically erased without any feedback. The 10 test problems were designed to have different stick lengths so they would span the range of relative length cues. Specifically, there

Table 1
Proportion of Problem Types in Each of the Four Conditions for Experiment 1

	Proportion of problems of each type			
Condition	Undershoot closer undershoot solved	Undershoot closer overshoot solved	Overshoot closer overshoot solved	Overshoot closer undershoot solved
Not biased/not predictive	.25	.25	.25	.25
Not biased/predictive	.40	.10	.40	.10
Biased/not predictive	.35	.15	.15	.35
Biased/predictive[a]	.48	.02	.30	.20

[a]These proportions lead to 68% undershoot-solved problems and 78% cue-predictive problems (as opposed to 70% and 80% as specified in the design). This slight difference was allowed so that participants in this condition would encounter each problem type at least once during training.

were two problems with a strong cue to choose undershoot (i.e., Stick C was much closer to the desired length than Stick B); two problems with a weak cue to choose undershoot (i.e., Stick C was somewhat closer to the desired length than Stick B); two neutral problems (i.e., Stick C and Stick B were equally close to the desired length); two problems with a weak cue to use overshoot (i.e., Stick B was somewhat closer to the desired length than Stick C); and two problems with a strong cue to use overshoot (i.e., Stick B was much closer to the desired length than Stick C). The strong-cue problems were most similar to those presented during training, although none of the test problems was identical to any of the training problems.

At the end of the experiment, participants were asked several questions to probe their task representations and strategies (e.g., "How did you decide which stick to use in your first solution step?" "Did you use any particular strategies to solve these problems?").

Analysis. The main dependent measure of interest was the solution procedure participants chose for each test problem. To identify the chosen solution procedure, we categorized participants' first stick selection as either longer than the desired stick (an overshoot choice) or shorter than the desired stick (an undershoot choice). Participants almost never chose the smallest building stick in their first move, so our categorization essentially distinguished between the choice of the long versus medium-sized building stick. In the following analyses, we present these choice data in terms of the percentage of trials for which the choice corresponded to the participants' more successful procedure. In the not-biased base rate conditions, in which both procedures were equally successful, this "more successful" label is arbitrary. Thus, we randomly assigned undershoot and overshoot to this label for participants in the not-biased conditions.

Results and Discussion

Prediction 1: Salient features of the task will influence initial cue use. The first prediction was that participants would tend to represent the salient features of the task and hence generate strategies that would make use of those features. Because these strategies are the basis for action, participants' initial choice behavior should be sensitive to the features they initially represent. In the BST, we focused on the relative length cue as a salient feature pattern that participants would likely include in their task representation. This suggests that on the test problems before training, participants' choices should exhibit sensitivity to this cue (i.e., there should be differences in choice proportions across the five test problem types that vary in relative lengths). Moreover, the RCCL model claims that salient features are combined into strategies in a way that is consistent with prior knowledge. Thus, according to our more specific prediction, we would find that participants would exhibit sensitivity to the relative length cue in a way that is consistent with the general notion of hill climbing (i.e., they would prefer the undershoot procedure on undershoot-closer problems and the overshoot procedure on overshoot-closer problems).

Figure 4 shows participants' choice tendencies for the five test problem types in both the initial and final tests. The four panels show the results for the four conditions. With respect to Prediction 1, we focused on the curve corresponding to the initial test in each panel. Note that all four conditions show a similar initial sensitivity to test problem type (i.e., all four initial test curves show an upward linear trend), $F(1, 296) = 246$, $MSE = 0.1$, $p < .001$. This suggests that participants represented BST features strongly related to the length cue and that these represented features played a role in their strategies for choice. For example, a reasonable description of this initial choice behavior is as follows: (a) If Stick B is closer to the goal than Stick C, then choose Stick B (begin overshoot). (b) If Stick C is closer to the goal than Stick B, then choose Stick C (begin undershoot). According to RCCL, these two if–then statements correspond to strategies that participants might generate based on their representation of the relative length cue.

Prediction 2: Under a given task representation, people will learn to prefer the strategies that have higher base rates of success. Prediction 2 states that participants will learn to prefer strategies that have been more successful in their past experience. In this experiment, as long as participants generated strategies that distinguish between the undershoot and overshoot procedures, they should learn to prefer the strategy corresponding to the more successful procedure. Figure 5 shows participants' choices between the two procedures, averaged by condition and by test. Here, we focused on choice at the final test because it reflected what participants had learned during training. Note that the *y*-axis measures the proportion of test trials on which participants chose the more successful procedure. For the biased conditions, this maps onto undershoot for half the participants and overshoot for the other half; the mapping is arbitrary for the not-biased conditions in which the two procedures were manipulated to be equally successful. An analysis of variance (ANOVA) on these data, with base rates and cue predictiveness as between-subjects factors, revealed a main effect of base-rate condition, with the biased base-rate participants choosing the more successful procedure more often than the not-biased base-rate participants, $F(1, 54) = 19.5$, $MSE = 0.22$, $p < .01$. The particular values of these aggregate choice results not only show base-rate sensitivity but are consistent with probability matching, a phenomenon found in other direct-experience situations (e.g., Estes, 1964; Gluck & Bower, 1988; Lovett & Anderson, 1995). Specifically, participants in the biased/predictive and biased/not-predictive conditions, who had 70% of their problems solved by the more successful procedure, chose that procedure on 72% and 63% of the test problems, respectively. Solvers in the not-biased/predictive and not-biased/not-predictive conditions, who had half of their problems solved by each procedure, chose the more successful procedure on 45% and 47% of the test problems, respectively.

This difference among conditions at the final test was not due to differences before training; there were no differences at initial testing because of base-rate condition ($F < 1$) or because of predictiveness condition ($F < 1$). Moreover, a mixed ANOVA including the two between-subjects factors and the within-subjects factor test (initial or final) indicated a greater shift in choice proportions from initial to final test among the biased conditions, $F(1, 72) = 4.94$, $MSE = 0.21$, $p < .05$. These results indicate that participants in the different conditions started out with fairly neutral choice tendencies regarding the two procedures and that after

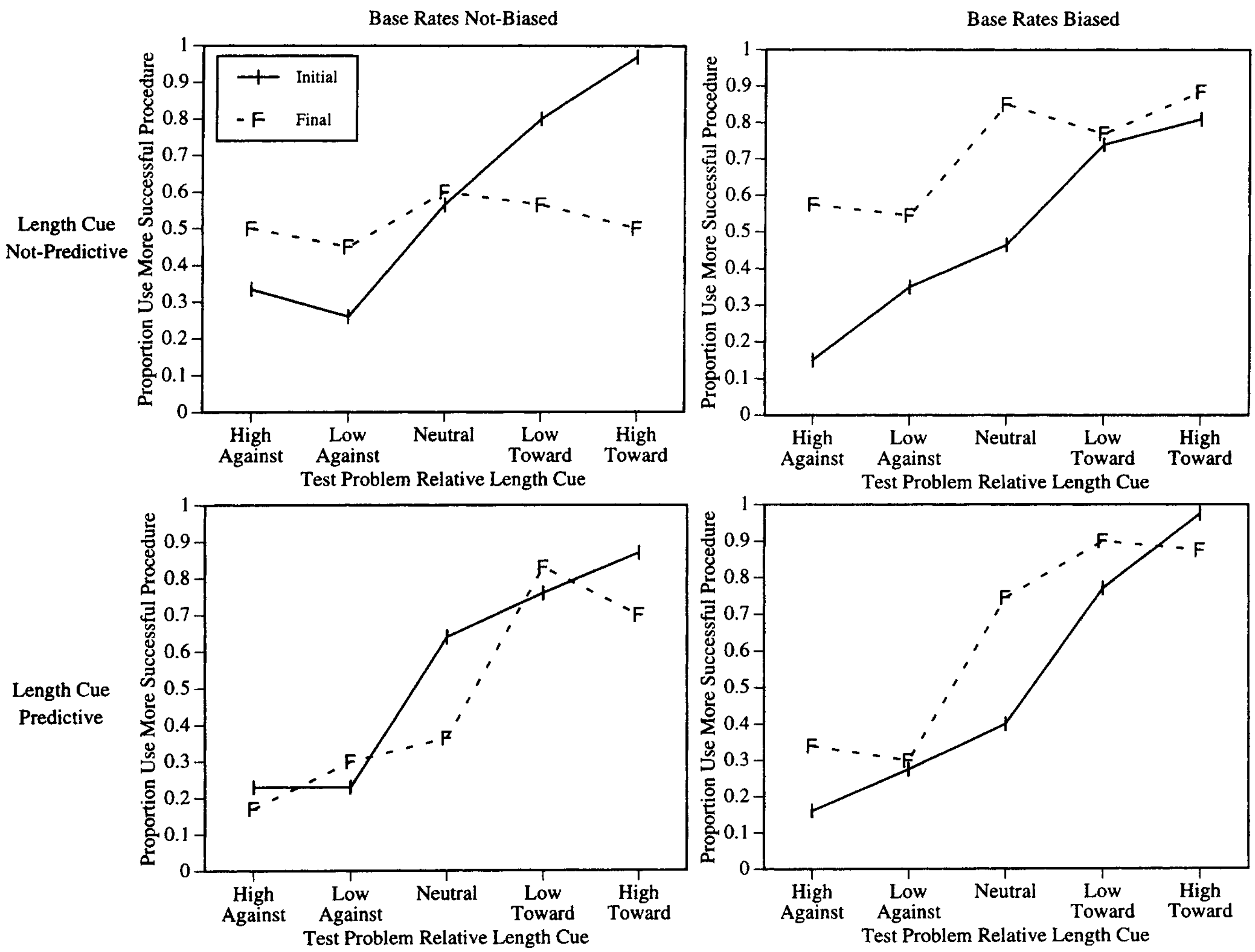

Figure 4. Choice proportions as a function of test problem type in Experiment 1 for the initial and final test phases. Relative length cue "high toward" refers to the test problems with relative stick lengths that made the overall more successful procedure appear closer to the goal. "Low against" refers to the test problems with a weak relative length cue that made the overall more successful procedure look farther from the goal.

training they did exhibit choice preferences consistent with the base rates corresponding to their condition, as the RCCL model predicted. The participants were able to show base-rate sensitivity because their initial task representation included a feature that distinguished between the overshoot and undershoot procedures.

Prediction 3: People will drop cues that prove to be irrelevant. The third prediction was that participants should drop features from their representation when the strategies derived from that representation are not successful. Here, we have evidence that participants initially included features related to the relative length cue in their representation (see Figure 4, initial test curves). Basing one's choices on this cue, however, will not be successful in the not-predictive conditions. Therefore, the RCCL model predicts that, by the final test phase, participants in the not-predictive conditions would be less sensitive to this cue than would participants in the predictive conditions. We explored this issue by investigating participants' explicit reports at the end of the experiment and by studying their final choice behavior across the five test problem types.

Explicit reports. At the end of the experiment, participants were asked to describe how they went about solving the BST problems. In particular, they were asked how they decided which building stick to use for their first move. This question was designed to tap participants' explicit access to their task representation and strategies. We categorized the participants' free-form responses into three major groups: those indicating sensitivity to various stick lengths, those indicating exclusive choice of a single procedure, and a miscellaneous group. This categorization was done by key words. If the response mentioned stick lengths, it was coded as a length-sensitive strategy. Many of these length-sensitive responses made comparisons between Stick B and the

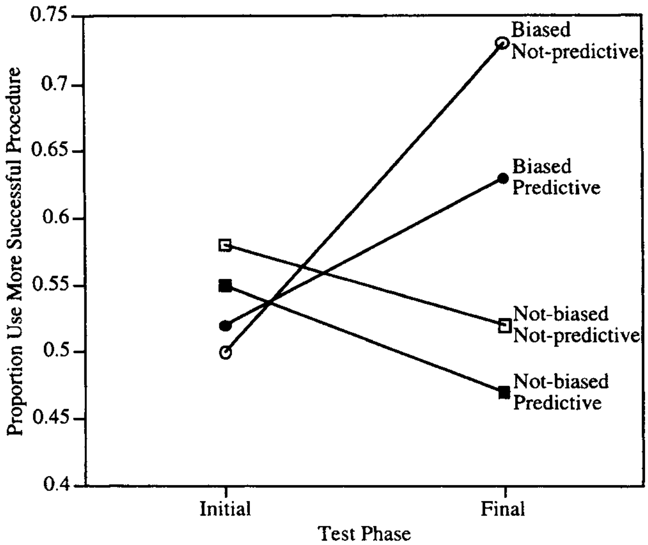

Figure 5. Base-rate sensitivity for Experiment 1 by condition for initial and final tests.

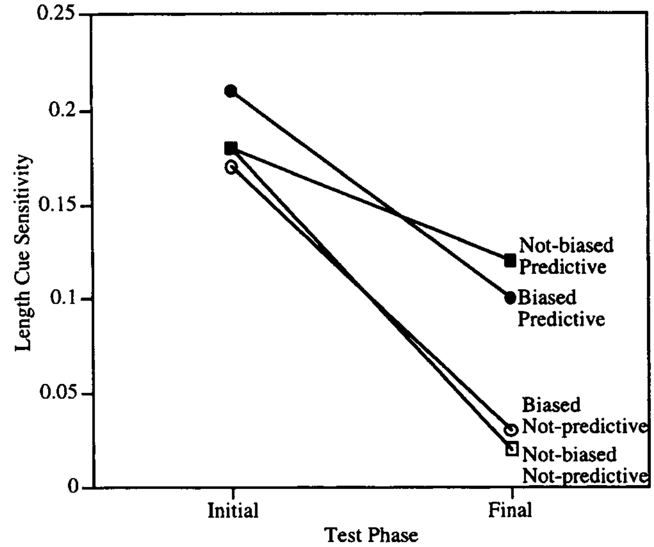

Figure 6. By condition in Experiment 1, initial and final sensitivity to the relative length cue, measured as the slope of choice curve across five test problem types.

desired stick or Stick C and the desired stick, similar to our definition of the relative length cue, which we hypothesized participants would naturally represent. If the response indicated that the participant had used a single procedure "always," "usually," or "generally," it was coded as an exclusive strategy. The miscellaneous category included all other responses, which often appeared to be guessing strategies. Table 2 shows the percentage of participants in each condition whose responses fell into each category. Our prediction that participants would be less likely to use the length-based cue in the not-predictive conditions relative to the predictive conditions was supported by these explicit reports: Not-predictive conditions had 12% length responses and predictive conditions had 40% length responses ($z = 4.8$, $p < .001$).

These explicit reports are our first evidence that participants in different conditions differed by the end of the experiment in their representations of the task.[1] Analyses comparing these explicit reports with participants' behaviors during training support the reports' validity: Solvers who reported "exclusive" strategies chose the more successful procedure more often than did solvers reporting other strategies, $F(1, 55) = 3.2$, $MSE = 0.07$, $p < .05$, one-tailed. Furthermore, participants who reported using stick lengths showed more sensitivity to the relative length cue than did participants reporting the other strategies, $F(1, 55) = 3.1$, $MSE = 0.43$, $p < .05$, one-tailed, when sensitivity was measured as the difference in choice preference between undershoot-closer and overshoot-closer training problems. Thus, the participants did appear to make explicit changes in the task representations and these changes were related to performance differences.

Table 2
Percentage of Participants Reporting Certain Strategies in Experiment 1

	Strategy report		
Condition	Length	Exclusive	Other
Not biased/not predictive	13	27	60
Not biased/predictive	30	40	30
Biased/not predictive	11	78	11
Biased/predictive	50	50	0

Behavior data. Regarding the behavior data, the RCCL model's prediction is that on the final test, participants in the not-predictive conditions will show less sensitivity to the relative length cue than will participants in the predictive conditions. Referring to Figure 4, one sees that the conditions did differ in their final sensitivity to the test problem's relative length cue. An ANOVA on these final data alone (using base rates and cue predictiveness as between-subjects factors and test problem type as a within-subjects factor) revealed the expected interaction between predictiveness and test problem type, $F(4, 288) = 3.34$, $MSE = 0.105$, $p < .05$. Specifically, there was less of an effect of test problem type in the not-predictive conditions than in the predictive conditions. This suggests that, at the end of the experiment, the not-predictive conditions were less sensitive to the relative length cue than were the predictive conditions.

Participants' degree of sensitivity to the relative length cue is perhaps best quantified by computing the slope of their choice proportions against test problem type (e.g., the slope of the curves in Figure 4). Figure 6 shows these slope

[1] Note that the explicit reports also show a fairly high level of within-conditions, between-subjects variability. This is entirely consistent with the model's predictions of task representation change following feedback.

data, averaged by condition and test.[2] As predicted by the RCCL model, at the final test, the average slope for the not-predictive conditions was flatter than that for the predictive conditions, $F(1, 56) = 10.08$, $MSE = 0.02$, $p < .01$. Indeed, the former was not significantly different from zero, $t(27) = 0.2$, *ns*, indicating no difference in choice tendencies across the different test problem types. Moreover, these data emphasize each condition's change in sensitivity from the initial test to the final test. The finding that all conditions showed a decrease in slope, $F(1, 54) = 6.93$, $MSE = 0.01$, $p < .05$, probably reflects the fact that even in the predictive conditions, the relative length cue was only 80% predictive. However, the most striking slope changes between initial and final test occurred for the not-predictive conditions, $F(1, 54) = 4.43$, $MSE = 0.01$, $p < .05$.

In summary, these results indicate that some representational change occurred among all conditions but that the greatest change occurred in the not-predictive conditions. It was in these conditions where the initially preferred strategy of choosing the stick that gets closest to the goal would not be successful. Thus, consistent with the RCCL model, participants in the not-predictive conditions stopped using the cue that played a large role in their original representations, as reflected in both their verbal and behavioral data.

Prediction 4: More representation and strategy change will occur in tasks with low success. Prediction 4 involves the relative degree of change that would be expected to arise under different task conditions. In particular, the RCCL model proposes that solvers will change their representation (and strategies) when their current strategies are fairly unsuccessful rather than uniformly and stochastically over time. The condition in this experiment in which any set of strategies would have limited success is the not-biased/not-predictive condition. This is because neither a strategy based on base rates of undershoot versus overshoot nor a strategy based on the relative length cue (nor any other strategy) would lead to success in this condition. The overall success rate of this condition attests to its difficulty: Participants in the not-biased/not-predictive condition solved 52% of problems in the final training block within the first five moves, whereas the other conditions averaged 60% (biased/not-predictive), 63% (not-biased/predictive), and 58% (biased/predictive), $F(1, 54) = 5.85$, $MSE = 0.014$, $p < .05$. The specific RCCL prediction, then, is that more change should arise in this most difficult condition: not-biased/not-predictive.

To quantify representational change and to demonstrate that it occurs to varying degrees at the individual level, Figure 7 (bottom) shows 6 participants' choice patterns across the training trials. Each panel presents a two-dimensional matrix of cells for characterizing an individual's sensitivity to the relative length cue (along the x-axis within a matrix) and base-rate sensitivity (along the y-axis within a matrix, measured as the proportion of choice of more successful strategy). The top matrix in Figure 7 provides a template for displaying the different behaviors that the matrix can describe. For example, the bottom-right cell (Cell i) represents choice behavior that shows positive sensitivity to the relative length cue and a tendency to choose the more successful procedure only rarely (approximately 10% of the time). In contrast, the top center cell (Cell b) reflects a strong tendency to choose the more successful strategy but no sensitivity to stick lengths.

In the bottom six panels, the Numbers 1–4 in each matrix indicate which cell best characterizes the participant's choice behavior on the first through fourth training block during the experiment. The path from Data Point 1 to Data Point 4 in a given panel shows the changes in choice tendencies of that individual across time. Participants with data points all in the right half of the two-dimensional grid maintained a sensitivity to stick lengths throughout the experiment, whereas participants with data points varying along the x-axis changed their sensitivity. Similarly, participants with data points at a constant height in the grid maintained a certain preference between the two procedures, whereas participants with data points at varying heights changed their preference. These different trajectories highlight the different patterns of change exhibited by participants in this experiment. For example, Participant 204 showed no change over the course of the task. This is not surprising because this participant was in the not-biased/predictive condition, which likely matched people's initial expectations of the task. In contrast, Participant 895 showed large strategy changes over the course of the task, varying both whether one strategy was preferred and whether the relative length cue was used. Again, this was not surprising given that this participant was in the not-biased/not-predictive condition, for which no strategy would be successful.

This matrix representation makes it possible to compute how "far" an individual has moved in this two-dimensional strategy space through the course of the experiment. For each participant, we computed the city-block movement across the four training blocks. For example, Participant 204 in Figure 7 had a city-block distance of zero, and Participant 895 had a city-block distance of 7 (3 + 2 + 2). The participants in Figure 7 were chosen to represent the 0th, 20th, 40th, 60th, 80th, and 100th percentile in terms of this distance metric. Using this city-block distance, we compared the amount of change in the not-biased/not-predictive condition with the other three. The mean city-block distance for participants in the not-biased/not-predictive condition was higher (3.3 units) than that of the participants in the other conditions (2.4 units), $F(1, 54) = 2.9$, $MSE = 3.0$, $p < .05$, one-tailed. Thus, participants in the most difficult condition appeared to change their task representation and strategies more over the course of the experiment than did participants in the "easier" conditions. However, the success rates were

[2] These data were computed by submitting each individual's initial and final test data to a linear regression and estimating the slope coefficient. The independent variable in each regression was the test problem's relative length cue (coded 1–5 to represent a continuum from "strong cue against the more successful procedure" to "strong cue toward the more successful procedure"), and the dependent variable was the proportion of test problems on which the more successful procedure was chosen.

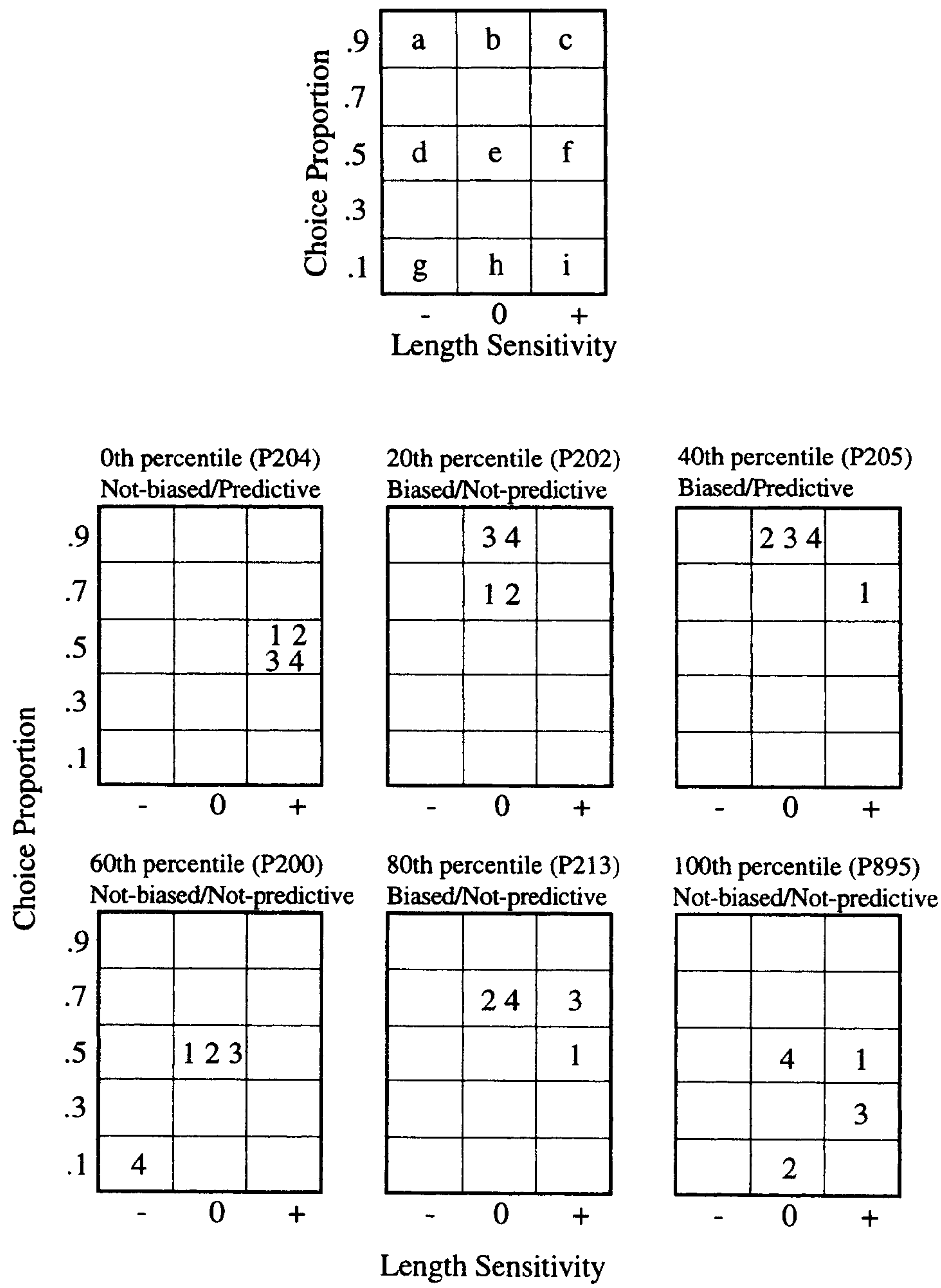

Figure 7. Top: Two-dimensional grid for describing choice behavior in the Building Sticks Task. Each cell characterizes an individual's choice behavior according to his or her choice of the more successful strategy and sensitivity to the relative length cue. Bottom: Individual choice pattern trajectories for 6 participants in Experiment 1. Numbers 1–4 indicate categorized behavior in the first through fourth blocks of training trials for each participant.

not high in any of the conditions, resulting in generally high levels of change.

Summary

Given these results, participants appeared to make their choices according to a well-balanced combination of base-rate information and problem-specific (cue) information. Participants' overall tendency to choose the more successful procedure shifted upward (between the initial and final tests) for the biased base-rate conditions but stayed the same for the not-biased base-rate conditions. In addition, participants' sensitivity to the relative length cue, although strong for all conditions initially, was maintained in the predictive conditions and was substantially reduced in the not-predictive conditions. Thus, participants showed reasonable choice behavior: Their choices were sensitive to a source of information only when that source of information reliably led to success.

Experiment 1 also produced four specific findings, each of which was predicted by the RCCL model. First, we found that participants initially made their choices in a way that was sensitive to the relative length cue. The RCCL model predicted this because the features making up this cue were salient and, based on prior knowledge, highly relevant to solving BST problems. Second, we found that participants were base-rate sensitive, choosing the more successful procedure more often in the biased base-rate conditions. RCCL predicts that people will be base-rate sensitive as long as their constructed strategies consistently distinguish among the possible choices. Third, we found that the task representations and strategies that participants used could be influenced by manipulating the predictiveness of certain task features; in particular, participants in the not-predictive conditions reduced their use of the relative length cue. This outcome was predicted because the RCCL model posits that people should drop features from their task representation (and incorporate new ones) when their current features and strategies are not leading to success. Fourth, we found that the degree to which individuals changed their choice behavior was related to the lack of success of their condition. This result was predicted by RCCL because having less successful strategies in general should lead to more representational change.

All of Experiment 1's results were found in the context of a problem-solving task, the BST, in which it is intuitively reasonable that participants generate and use various strategies for making their problem-solving choices. In other choice tasks, the role of problem-solving-like strategies is less obvious. Therefore, in Experiment 2, we explored related issues regarding base-rate and cue sensitivity in the context of a simpler categorization task.

Experiment 2: The Colors Task

Experiment 2 was focused on base-rate and cue sensitivity in the Colors Task. This task was chosen because it allows the study of representation and strategy change in a choice context where we could manipulate participants' representations not only by varying the probabilities of different outcomes but also by varying the overall "look" of the task. In particular, we designed two distinct versions of the task so that participants would approach the two versions with different task representations. In this way, not only should participants in different base-rate conditions *learn* to represent the task differently on the basis of their experience (as they did in Experiment 1), but those given different task versions should also *initiate* the experiment with different representations and strategies. In particular, the goal of Experiment 2 was to test whether different initial representations of the same underlying task could lead to different base-rate effects. In Experiment 2, we also extended our use of explicit reports as an additional means of gathering information about participants' task representations and strategies.

The other reason for choosing the Colors Task is that it is the only case in the literature that we know of that claims to find base-rate neglect on training trials in a direct-experience situation (Goodie & Fantino, 1995). We wanted to explore the task that had led to this unusual result. Essentially, the original Colors Task (Goodie & Fantino, 1995) is a delayed match-to-sample task in which the correct choices are determined by fixed probabilities. Each trial begins with the presentation of a cue (also called the "sample"), a single rectangle colored either blue or green. After a delay, two rectangles appear, one in the color of the cue and one in the other color. The participant then chooses one of these two rectangles (i.e., either the one matching the cue color or the other one) and receives feedback about whether the choice was correct. Note that the probability of being correct by matching to the blue cue can be different from the probability of being correct by matching to the green cue. Goodie and Fantino manipulated these two probabilities across conditions. They found that, even after approximately 200 trials of practice and even with monetary payoff as an incentive, participants' choices did not always conform to the cue-specific probabilities of their condition. For example, in one condition in which matching the cue color was correct 80% of the time for one cue color and 50% of the time for the other cue color, participants' choices averaged 78% and 70% matching those two colors, respectively. Similarly, in another condition in which matching the cue color was correct 67% of the time for one cue color and 33% of the time for the other cue color, participants' cue-matching choices averaged 62% and 56% for the two cue colors, respectively. In each condition, participants' matching percentages were highly similar for the two cue colors (less than 10% apart) despite a substantial difference between the corresponding percentages produced by the task (approximately 30% apart). This led Goodie and Fantino to conclude that participants were not adequately sensitive to base rates, specifically the base rates of the two choice colors. One can see that this is true from the description of their data: Matching the two cue colors approximately equally often implies that participants were choosing the two choice colors approximately equally often, yet the matching probabilities for the two cue colors were different, implying one choice color was correct more often than the other.

Note, however, that participants in Goodie and Fantino's (1995) experiment were sensitive to another base rate, the overall base rates of success of matching the cue color. In the first condition, in which matching the cue color was correct on average for 67% of all trials, participants' average matching rate across the two cue types was 74%, and in the second condition, in which matching the cue color was correct on 50% of all trials, participants' average matching rate was 59%. That is, participants were more likely to choose to match the cue color in the condition for which this was more successful. Thus, even in Goodie and Fantino's experiment, contrary to their claims, people are base-rate sensitive in this experiential task; they are sensitive to the base rates of matching versus not matching the cue color. Even from this "matching" perspective, however, the participants in Goodie and Fantino's experiment displayed an interesting departure from normative behavior: They did not adequately integrate cue-specific information with overall base rates. That is, they were not sensitive to the separate success rates for matching to each cue color. We call this kind of sensitivity (or lack thereof) *cue-specific base-rate sensitivity.*

The RCCL model offers a representation-based account of these results. According to RCCL, the key to overall and cue-specific base-rate sensitivity depends on having the relevant features in one's task representation. If participants include in their representations and strategies the "match/not-match" status of each trial but not the specific colors, then they will be sensitive only to the overall base rates of matching. In contrast, if participants include in their task representations and strategies the specific colors for each trial, then they should be able to learn the different base rates associated with each cue color and exhibit case-specific base-rate sensitivity.

We tested this explanation in Experiment 2 by varying the superficial features of the task to make the individual cue colors more salient and the "match" feature less salient. This manipulation, according to the RCCL model, should produce stronger cue-specific base-rate sensitivity. More important, this manipulation serves as a more direct test of the RCCL model's predictions regarding the role of task representations in initial choice behavior, as compared with the tests in Experiment 1.

The overall structure of Experiment 2 was to study base-rate sensitivity in the original Colors Task as well as in our own modified version of it. The original version of the Colors Task is referred to as "2 colors." Our modified version ("4 colors") maintains the same structure and probabilities of the original Colors Task but uses four colors in total: The cue can take one of two colors, as in the original version, but the choices are two additional colors.

Given that participants will tend to encode cue and choice colors that are the same as "matching," participants in the "2 colors" task may represent the choice colors as "matching" or "not matching" the cue color, whereas participants in the "4 colors" task will not be able to do so (i.e., the choice colors are never the same as cue colors in the "4 colors" task). With such differing representations, the RCCL model predicts that the two task versions evoke different strategies: "2 colors" participants will generate strategies that involve matching the cue color, whereas "4 colors" participants will not. Because the RCCL model claims that people learn from their experience in a task in terms of the choice strategies they are using, it predicts that the difference in strategies generated under the two task versions will lead to different patterns of learning.

In summary, the RCCL model makes three key predictions for Experiment 2:

1. *Overall base-rate sensitivity depends on strategy sets.* If participants represent the task in terms of match and not match, then they should be sensitive to the global base rate of success of match and not match. Because the participants in the two-color condition are expected to represent the task in this fashion, they are expected to choose to match more often in the 80/50 condition (67% match success rate overall) than in the 67/33 condition (50% match success rate overall). In contrast, the participants in the four-color condition are not likely to represent the task in terms of match and not match and therefore should not be sensitive to the overall base rate of match and not match (i.e., they will show little or no difference in overall match rates across the 67/33 and 80/50 conditions).

2. *Cue-specific base-rate sensitivity depends on cue saliency.* Because the four-color participants are expected to represent the separate cue colors, they should be able to learn the separate base rates associated with each cue color. In contrast, because the two-color participants are expected to abstract across the two cue colors and represent only the choice colors' match status, they should be less able to learn a separate base rate associated with each cue color.

3. *Changes in task representations will be a function of task difficulty.* Not only does the RCCL model predict asymptotic choice tendencies within each of the conditions, but it also predicts that participants can change their task representations (and corresponding strategies) during the task and predicts when these changes are more likely to occur. Specifically, as in Experiment 1, the model predicts that the degree of representation change over time will be a function of overall strategy success rates (i.e., condition difficulty). In Experiment 2, the two-color 67/33 condition is expected to be the most difficult condition because the task representation that is most obvious in that condition (match and not match) will be unsuccessful (i.e., at chance). Therefore, the RCCL model predicts most representational change in that condition.

Prediction 1 contrasts with an implicit memory view in that we are predicting that participants will learn only the base rates of features that they are explicitly representing. That is, even if other features occur with different base rates, participants will be able to learn only different base rates for the features included in their task representations. Implicit memory theories allow for base-rate learning to occur without explicit awareness. Prediction 1 also contrasts with Goodie and Fantino's (1995) suggestion that people cannot learn base rates in this experiential task. Instead, our model claims that people are always learning and using base rates from their experiences of success and failure; it is just that the strategies on which those base rates are being computed may not conform to the experimenter's intended task representation (e.g., measuring sensitivity to different base rates of choice color will not reveal people's sensitivity to different base rates of match vs. not match). Prediction 2, like Prediction 1 of Experiment 1, is consistent with other feature-weighting categorization models: The more salient a feature, the more likely it is to be influential in participants' choices (e.g., Anderson, 1991; Goldstone et al., 1997; Kruschke, 1996; Nosofsky, 1984). However, these other models do not make all the other predictions that the RCCL model does. In particular, Prediction 3 is especially novel: RCCL predicts that representational change will occur over time even in a simple choice task and that such change is more likely in certain circumstances. These three predictions were tested by examining participants' choice data and their explicit strategy reports.

Method

Participants. Eighty Carnegie Mellon undergraduates participated for course credit and were randomly assigned to one of four conditions.

Design. There were four experimental conditions that differed according to two factors: (a) the version of the Colors Task ("2 colors" vs. "4 colors") and (b) the probabilities determining correct choices "67/33" and "80/50." These probability pairs are the same as those used by Goodie and Fantino (1995). Figure 8 shows the cue–choice combinations to which these probabilities apply (both for "2 colors" and "4 colors"). For the "67/33 two-color condition," the probability that choice Color A would be correct after Cue A was .67, and the probability that choice Color B would be correct after Cue B was .33. For the "80/50 two-color condition," the probability that choice Color A would be correct after Cue A was .80, and the probability that choice Color B would be correct after Cue B was .50. Thus, in the two-color case, the probability label indicates the probability that the correct choice matched the color of the cue for Cue Colors A and B. In both conditions, A is called the more reliable cue and B the less reliable cue because Cue A is a more reliable predictor of the matching response. The probabilities were the same for the four-color conditions, but here the cue and choice colors were never the same, so color "matching" refers to the choice color that is paired with each cue color, by analogy to the two-color conditions. Note that for each participant, either two or four colors were chosen from the set {red, blue, yellow, green} and randomly assigned to the various cue and choice colors as required for that condition.

Procedure. At the beginning of the experiment, a computer tutorial provided participants with instructions and practice on a black-and-white "2 colors" version of the Colors Task. Each participant was then given 200 trials. Each trial was initiated with a "ready?" message that disappeared as soon as the participant clicked the computer mouse. Then, a single square-shaped cue color would appear. The cue would remain on the screen until the participant clicked on it with the computer mouse. Then, after the screen remained blank for 3 s, the two choice colors would appear in square shapes. The assignment of each choice color to left–right placement was random. The choice colors would remain on the screen until the participant clicked on one of them. Finally, the choice colors would disappear, and a feedback message would be given (both visually and aurally), indicating whether the choice was correct. The total number of points accumulated so far (+1 for correct, +0 for incorrect) was also displayed at this time. The feedback message remained on the screen until the participant clicked the computer mouse but for a minimum time of 1 s. Then, the next trial's "ready?" message would appear. Figure 9 shows an overview of the task structure.

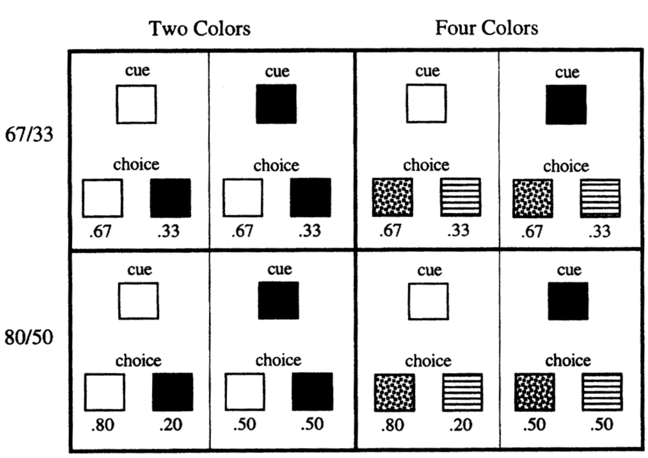

Figure 8. The probability of each choice being correct with each given cue for all four conditions (two color vs. four color × 67/33 vs. 80/50) of Experiment 2. Here the patterns are used to represent the different colors participants saw.

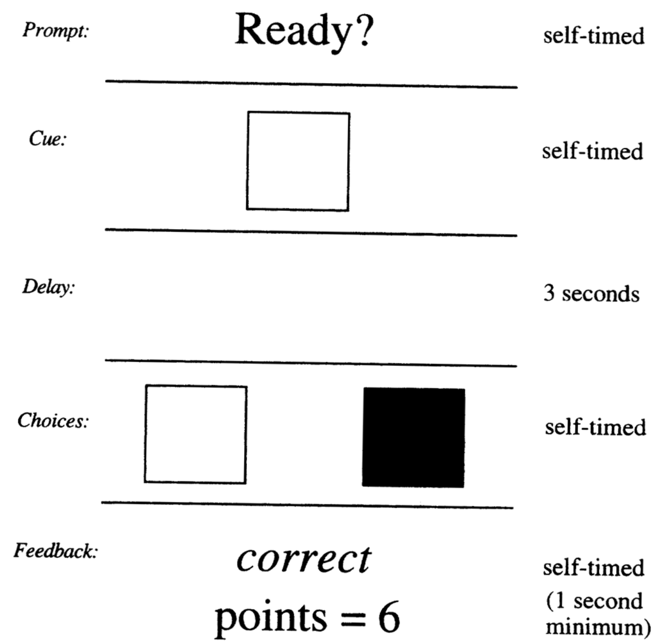

Figure 9. The trial structure of the Colors Task (Experiment 2). Note that participants saw colored squares rather than black and white squares.

Postexperimental questions were administered verbally by the experimenter. These questions asked the participant to describe in words the various choice strategies that he or she had used throughout the course of the experiment. The participant was asked to indicate which of these strategies had been used most recently (i.e., at the end of the experiment). These questions served to provide convergent evidence regarding the features that the participants represented, the strategies they used, and the degree of representational change over the course of the task. Additionally, for each strategy verbalized by a participant, the following specific questions were asked: "Overall, how often did you use the strategy?" "Of those times, how often was that strategy successful?"

Analysis. The main dependent measure in this experiment concerned the choice color that the participant chose on each trial as a function of the cue color presented on that trial. The analyses show these choice data in terms of whether the participant's choice matched the color of the cue, which is referred to as *color matching.* For the "2 colors" version of the task, this dependent measure is intuitive and follows the reporting style in Goodie and Fantino (1995): A choice matched the cue if it was the same color. For the "4 colors" version of the task, choices could never be the

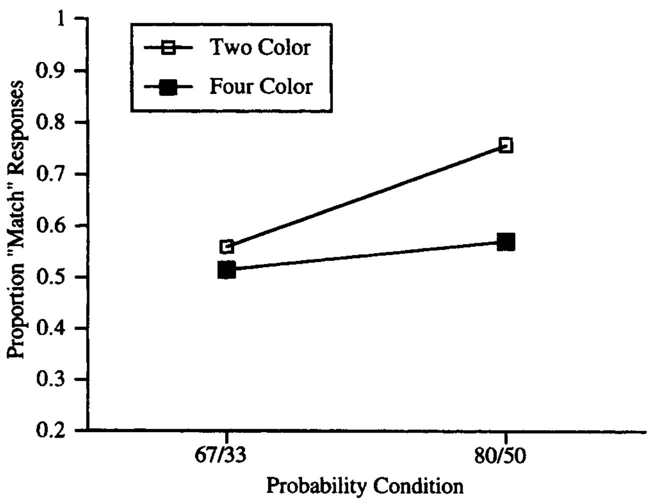

Figure 10. The mean proportion of match responses by condition (Experiment 2).

same as the cues, but there was a mapping between the two choice colors and the two cue colors based on the correspondence with the "2 colors" task. Therefore, we consider cue and choice colors as "matching" in the "4 colors" task if the corresponding colors would be truly matching colors in the "2 colors" task.

Results and Discussion

Prediction 1: Overall base-rate sensitivity depends on strategy sets. The first prediction is that overall base-rate sensitivity in the degree of matching responses will occur only when participants use strategies that represent the match/not-match feature. That is, the participants in the 80/50 conditions (67% overall matching) should have higher matching rates than the participants in the 67/33 conditions (50% overall matching), but only in the two-color versions. Participants in the four-color versions should not show differences in overall matching rates between the 80/50 and 67/33 conditions because they would tend not to represent the task in terms of matching and so would be learning the success of strategies based on other features. Figure 10 shows the mean proportion of match responses within each of the four conditions across all trials. An ANOVA on these data, using the two between-subjects factors, showed the expected interaction of task version and probability condition, $F(1, 76) = 10.0$, $MSE = 0.01$, $p < .005$. Specifically, the effect of probability condition was strong for the two-color versions, $F(1, 38) = 38.5$, $MSE = 0.01$, $p < .0001$, with a 20% difference in matching rates between the two probability conditions. In contrast, there was only a 6% difference that was marginally significant for the four-color versions, $F(1, 38) = 3.0$, $MSE = 0.01$, $p < .1$.

Prediction 2: Cue-specific base-rate sensitivity depends on cue saliency. The second prediction is that sensitivity to the different base rates associated with each cue color will vary as a function of cue saliency. More specifically, because individual cue colors are less salient in the "2-colors" task, participants in the two-color conditions should exhibit a smaller difference in match rates between the more reliable and less reliable cue colors than participants in the four-color conditions. Figure 11 shows the percentage of trials on which participants' choice color matched the cue color for the more reliable and less reliable cue colors. An ANOVA using the two between-subjects factors and cue type as a within-subjects factor showed that the crucial interaction of cue type with task version was significant, $F(1, 76) = 4.1$, $MSE = 0.02$, $p < .05$, such that there was stronger sensitivity to cue type in the four-color version than in the two-color version. This result is consistent with the RCCL model's prediction that participants in the four-color version would be more likely to incorporate the specific cue and choice colors in their choice strategies and hence be more sensitive to different base rates for the two cue colors. The interaction of cue type with probability condition and the three-way interaction of task version, probability condition, and cue type were not significant (Fs < 1), demonstrating that this effect of task version on sensitivity to cue type did not vary across the probability conditions, as expected.

Another way of quantifying participants' cue-specific base-rate sensitivity involves computing the difference between each participant's color-matching percentages for the two cue types (more reliable [MR] minus less reliable [LR]), without regard for the absolute percentage values. These differences were .12 and .08 in the two-color 67/33 and 80/50 conditions and .20 and .20 in the four-color 67/33 and 80/50 conditions. These difference scores confirmed that four-color participants showed greater sensitivity to the different base rates of the two cues than did two-color participants, $F(1, 76) = 4.1$, $MSE = 0.04$, $p < .05$.

In summary, the manipulation of the task ("2 colors" vs. "4 colors") produced the expected effects: Participants were

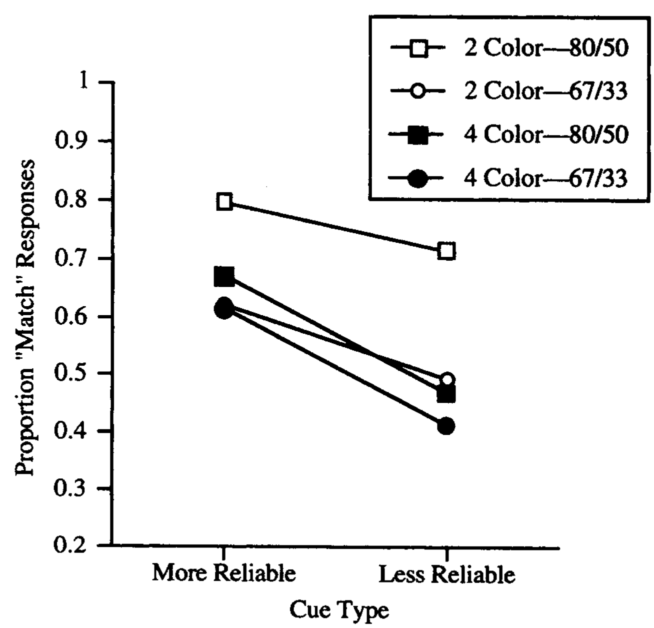

Figure 11. The mean proportion of match responses by cue reliability and condition (Experiment 2).

more sensitive to different cue-specific base rates in the four-color conditions than in the two-color conditions. It is not the case that these results could be attributed to participants in the four-color condition learning the probabilities more quickly such that their asymptotic performance contributed more strongly to averages taken over the entire experiment; these findings were maintained when the analyses were recomputed using the final 40 trials alone instead of all trials.

Prediction 3: Change in task representations will be a function of task difficulty. The third prediction involves the amount of within-subjects change in task representations and strategy use. To examine this issue, we developed a measure that combined an individual's probability of matching the more reliable cue color with the probability of matching the less reliable cue color. These two numbers can be represented in a two-dimensional display as in Figure 12 (top). Both dimensions can be partitioned into N regions (such as low, middle, and high probabilities in this figure), making $N \times N$ different cells (nine in this case). Some of these cells are labeled with a characterization of the behavior associated with that cell. For example, a high probability of matching both the more reliable and less reliable cue colors can be characterized as general color matching. A high probability of matching the more reliable cue color and a low probability of matching the less reliable cue color implies a fairly consistent choice of a single choice color, the choice color corresponding to the more reliable cue color. Nonlabeled cells represent behavior that is in between (possibly a mixture of) the two neighboring cells.

Figure 12 (bottom) shows 6 individual participants' paths through this two-dimensional choice space (on a 10×10 grid). Each panel locates a separate participant's choice behavior measured during Blocks 1–5. These participants were selected to represent all four conditions and to show the entire continuum of behavior change, from Participant S409, who spent all five blocks in the $\langle 1.0, 1.0 \rangle$ cell (i.e., matching the cue color exclusively),[3] to Participant S123, who appears to have bounced back and forth between relatively extreme matching behavior and more neutral behavior, reflecting changes between strategies that are sensitive to the cue color and strategies that are not. Each panel, taken alone, shows the change of a single individual across time.

Were these differences in within-subjects variability related to task difficulty, as the RCCL model predicts? The two-color 67/33 condition was the most difficult condition in that task structure biased these participants to adopt an initial representation and strategy set (i.e., match the cue color) that would be particularly unproductive, with a success rate of only 50%. Indeed, the participants in this condition had the lowest overall success rate (a mean of 51% vs. 54% for the other conditions), $F(1, 76) = 6.2$, $MSE = 25$, $p < .02$. According to the RCCL model, this low success rate should lead to the greatest representation and strategy change in the two-color 67/33 condition.

To measure within-subjects change, we counted the number of times the cell describing a participant's choice behavior (within the 10×10 grid) changed from one block of 40 trials to the next. For each participant, this produced a number between 0 (no change) and 4 (a change between each pair of adjacent blocks). Across all conditions, the participants moved a mean of 3.4 times from one block to the next (i.e., 85% of the time). As predicted, the participants in the two-color 67/33 condition had the highest number of block-to-block transitions in the 10×10 grid (a mean of 3.8 vs. 3.3 for the other conditions), $F(1, 76) = 4.1$, $MSE = 0.9$, $p < .05$.

The preceding analyses of changes in choice tendencies over time have been cast in terms of strategy changes. However, these analyses do not rule out the possibility that the participants used the same strategy throughout the task and simply shifted in their distribution of responses over time, either because of random variation or because of learning the relative success rates of each response. To examine this possibility and to provide additional tests of Prediction 3, we examined the participants' explicit self-reports.

At the end of the task, the participants were asked to report the strategies that they had used during the task. Thirteen identifiable strategies emerged from these self-reports, indicating that the participants did indeed use different strategies. Table 3 shows the list of strategies generated by the participants, ordered by their overall frequency.[4] The participants reported a mean of 3.8 strategies each, and several participants reported using as many as 7 different strategies. This large frequency of different strategies suggests that the participants did vary their strategies over time rather than simply using the same strategy.

How did the number of these strategies relate to condition difficulty? First, as a consistency check, the number of strategies reported correlated .30 ($p < .01$) with the number of moves in the 10×10 grid over time. This moderate but significant correlation was likely due to the fact that many of the strategies lead to behavior consistent with the "neutral" cell and thus any given strategy change would not necessarily result in a behavioral change according to our measures. Returning to the prediction relating strategy variability to condition difficulty, the participants in the most difficult condition (two-color 67/33) did have the highest number of reported strategies (a mean of 4.1 vs. 3.7 for the other conditions). Because all the conditions were fairly difficult, however, these differences were only marginally significant, $F(1, 76) = 1.9, MSE = 1.3, p < .2$.

Not only does the RCCL model predict that there will be high variability in task representation and strategy use when the overall success rates are low, but the RCCL model also predicts that strategies will be used to the degree to which they are successful (see Prediction 2 in Experiment 1).

[3] Participant S409 was the only participant who reported using only one strategy throughout the task: Always match the cue color.

[4] A subset of the data was recoded by a second coder given the 13 identified strategies. Interrater reliability for coding of the strategies was 96%. Also, the strategy identified by each participant as used more recently was consistent with choices made and latency patterns from the final trials of the experiment, indicating that the self-reports were accurate reflections of strategy use.

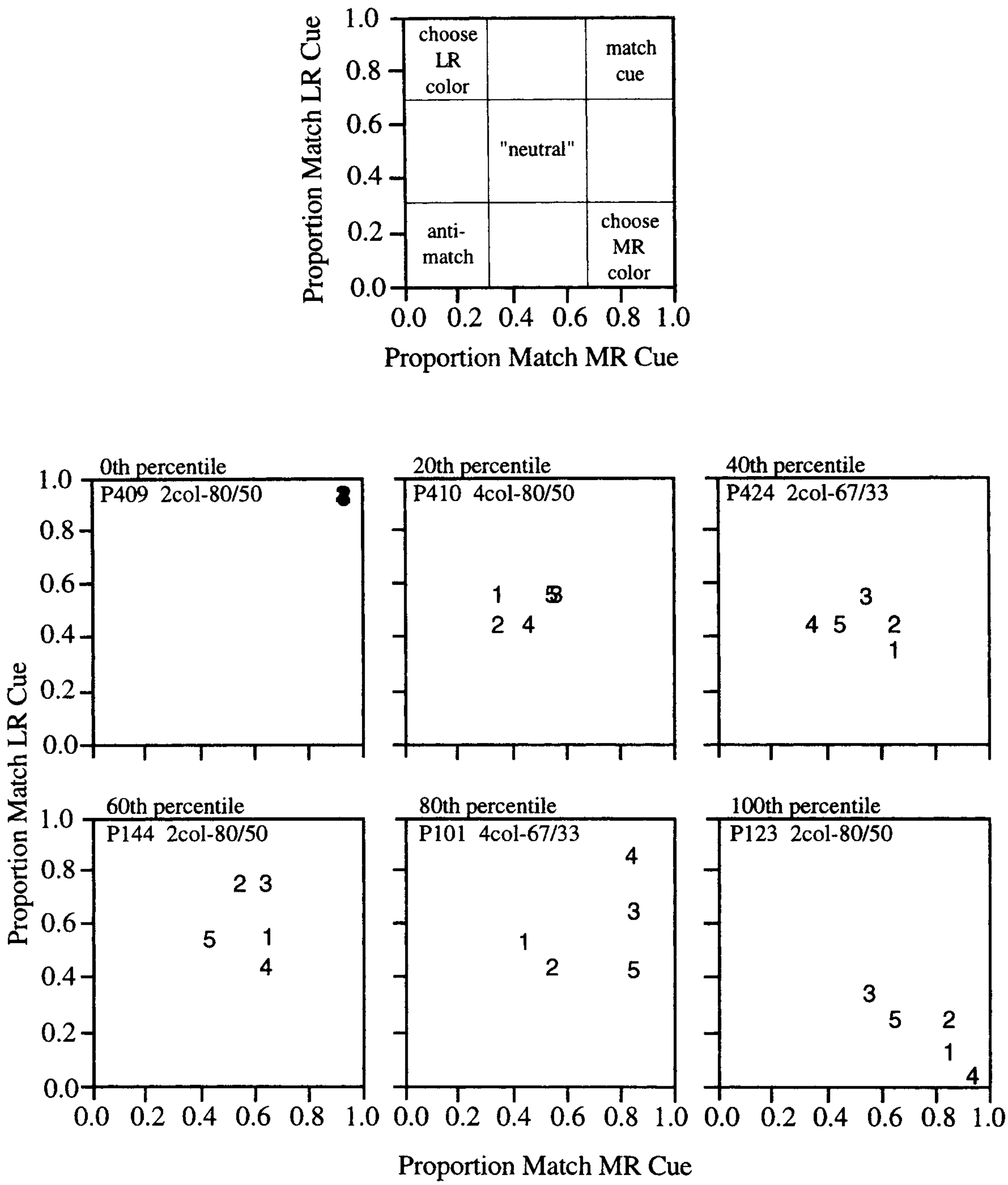

Figure 12. Top: Two-dimensional grid for describing choice behavior in the Colors Task. Each cell characterizes an individual's choice behavior according to his or her proportions for matching the less reliable (LR) cue and the more reliable (MR) cue. Bottom: Individual choice pattern trajectories for 6 participants in Experiment 2. Numbers 1–5 indicate categorized behavior in the first through fifth blocks of trials for each participant.

Although we do not have direct measures of strategy use and strategy success, the participants did provide their own estimates of how often they used each strategy and how often it was successful. As the RCCL model predicts, rated success was positively correlated with rated use ($r = .48$, $p < .0001$). Why was the correlation not closer to one? One important factor is that some of the more successful strategies may not have been discovered until later in the task. If this were the case, then participants would be much more likely to underuse a successful strategy relative to its measured success rate. Using a ±10% margin around perfect equality between a strategy's reported use and

Table 3
List of Strategies Generated by Participants With Proportion of Participants Using the Strategy as Their Final Strategy or at Any Point

Strategy	Last strategy	Overall
Cue matching	.14	.79
Choice color and side pattern	.18	.61
Complex choice color pattern	.18	.58
Random choice	.16	.54
Always one choice color	.15	.34
Complex side pattern	.04	.25
Always one side	.05	.18
Alternating choice colors	.03	.14
Alternating side	.05	.11
Antimatch	.00	.09
Success and failure pattern	.00	.06
Match and random	.03	.03
Location of click	.00	.01
Other	.01	.08

reported success, only 17% of the strategies were used more often than they were successful, whereas 34% of strategies were used less often than they were successful.

Summary

Experiment 2 produced several major findings that confirm predictions of the RCCL model. First, participants working on different versions of the task showed different degrees of base-rate sensitivity. At the global base-rate level, two-color participants were more sensitive to differences in global base rates of match and not match than were the four-color participants. Second, at the cue-specific level, four-color participants were more sensitive to the differential predictiveness of each cue color than the two-color participants. The RCCL model predicted these results because people's task representations for the two versions differed, leading the four-color participants to represent individual colors and two-color participants to represent the match/not-match feature (abstracting across the individual colors). It would be surprising if the participants were not encoding the colors of the cue squares in all cases (cf. Treisman & Gelade, 1980). This is not the prediction of the RCCL model, which predicts rather that certain participants (particularly those in the two-color condition) do not include color of the cue in their task representation and strategies, preventing them from learning the cue-specific success rates of matching. In contrast, an implicit memory account would have predicted that all participants would be able to learn these success rates, even those in the two-color condition. Third, we found evidence of between-subjects variability that was related to participants' self-reported strategy use, and we also found evidence of within-subjects variability that was related to participants' self-reported strategy change. The amount of variability also differed in ways predicted by the RCCL model: There was more variability in situations in which participants encountered less success. Moreover, the correlation between participants' reported strategy use and success rate confirmed the RCCL prediction that people will learn to choose strategies according to their rates of success. These latter findings replicate results from Experiment 1.

General Discussion

Summary

In this article we present a new model of the process by which people use their task representations to learn to make choices and of the process by which people change their task representation over time. This model makes predictions for when and how people integrate base-rate and cue information in making choices. In the two experiments reported here, we found evidence consistent with the main processes specified in the RCCL model:

1. Individuals represent the task at hand by selecting a subset of features to encode using feature saliency and background knowledge. In both experiments, participants varied across conditions in whether they exhibited sensitivity to certain task features, such as relative stick lengths (Experiment 1) and cue color (Experiment 2).

2. Various features in the task representation are combined to generate a particular set of strategies for choice. In both experiments, participants reported a variety of strategies, and their choice behavior was consistent with their reports.

3. Through experience, individuals learn and make choices according to the relative success rates of their strategies. In both experiments, the aggregate choice tendencies of each condition moved toward the strategies that would be most successful for that condition. Moreover, analyses by strategy revealed that participants selected strategies as a function of how often they thought they succeeded, producing overall base-rate sensitivity. However, as Experiment 2 showed, participants showed base-rate sensitivity only to the features that they included in their task representations. It is important to reiterate that this result could have gone otherwise; for example, implicit learning theories predict that people can learn associations, such as base-rate sensitivity, without explicit awareness. Instead, our model claims that people are always learning success base rates and using them to make choices, but it is the task representation that determines *which* base rates are learned and used.

4. When strategies in the current set are unsuccessful, people will tend to modify their task representation and generate new strategies. In both experiments, the tasks were quite difficult, and the majority of the participants changed their strategies over time. As the model predicted, participants in the more difficult conditions changed their choice behavior more than did participants in the less difficult conditions. Moreover, in Experiment 2, participants' number of reported strategies covaried with their amount of change.

A particularly novel aspect of the RCCL model is that it makes variability an integral component of choice. The RCCL model proposes that individuals may vary in their task representations and hence their strategies for choice and that a particular individual's set of choice strategies may change with time. According to the RCCL model, these are systematic and fundamental sources of variability. By performing individual participant analyses for both of the experiments, we aimed to demonstrate that these different types of variability do exist. There was consistency between the behavioral data and the explicit self-reports suggesting that the variability we found was neither spurious nor due to lapses of attention or boredom; rather, there were conceptual differences across participants and across time that would not have been revealed in aggregate analyses.

In both experiments, the success rates in all conditions were relatively low and, as we consequently argue, the levels of task representation change were relatively high in all conditions. However, note that the success rates in our experiments were not atypical of many other probabilistic categorization experiments (e.g., Busemeyer & Myung, 1992; Elliott & Anderson, 1995; Estes et al., 1989; Friedman et al., 1964; Gluck & Bower, 1988). Thus, we expect that similar levels of task representation change occurred in these previous experiments.

Although the model was tested in the context of two relatively simple experiential tasks for experimental control purposes, the simplicity of these tasks does not undermine the generality of the findings. That participants varied so much over time in the context of such simple tasks suggests that people always restrict the features that they use to represent a task and that their feature set will change over time. Given limited attentional capacity, more complex tasks should produce proportionally greater restrictions in the features used and greater feature variability over time. Indeed, related work using a much more complex scientific discovery task showed that people frequently changed their task representations from trial to trial (Schunn & Klahr, 1995, 1996).

The main processing stages of the RCCL model have been discussed loosely in terms of an information-processing model, in which strategies for choice are represented by if–then rules. The processes specified in the RCCL model can then be conceived of as mechanisms for generating new rules, learning the success rates of existing rules, and choosing from among competing rules according to their success rates. The model could easily be implemented in a specific computational model (e.g., using the ACT-R [adaptive control of thought—rational] theory; Anderson, 1993). Nevertheless, as it is described in this article, the RCCL model is more of a conceptual model that could be instantiated within a variety of architectures. This generic approach has the advantage of focusing our initial work on the novel aspects of the model—the roles of task representation and variability—and their relationship to largely unexplored choice phenomena without tying our claims to specific implementation details.

Related Findings

Other work by Goodie and Fantino (1996) showed results similar to those of our Experiment 2 with a modified version of the Colors Task that used a vertical or horizontal line instead of a blue- or green-colored rectangle for the cue. With these new stimuli and the same probabilities as in previous experiments, their participants demonstrated greater sensitivity to base rates, as we found in our four-color manipulation. They also found that administering this modified task to participants who were first trained to associate line direction with color (e.g., vertical with blue) showed a return to the initial task's base-rate neglect. From these results, then, Goodie and Fantino proposed that participants' prior knowledge of particular associations between cue and choice will affect their ability to exhibit cue-specific base-rate sensitivity (i.e., if there is such a preexisting association, then there will be no cue sensitivity). In contrast, we argue that the task representation and set of available strategies determines cue sensitivity.

Table 4
Predicted and Mean Observed Cue Sensitivity of Match Responses in the Last 20 Trials as a Function of Strategy Reportedly Used Last

Strategy	Predicted sensitivity	*n*	Cue sensitivity *M*	Cue sensitivity *SE*
Always one choice color	Strong	12	.67	.09
Choice color pattern	Moderate	14	.20	.08
Choice color and side pattern	Moderate	14	.14	.08
Side	None	4	.00	.15
Match cue	None	11	.01	.08
Random	None	13	.18	.06
Alternating side	None	4	−.12	.12

Note. Cue sensitivity was measured as matching proportion for the more reliable cue minus matching proportion for the less reliable cue.

One way to tease apart these theories is to examine the relationship between strategy use and cue sensitivity. Because the RCCL model casts cue sensitivity in terms of strategies used, it predicts that participants using strategies that are definitionally related to cue should show cue sensitivity in their choices, whereas participants using strategies that are definitionally unrelated to the cue should show no cue sensitivity in their choices. In contrast, Goodie and Fantino's account makes no such prediction.

Table 4 shows participants' mean cue sensitivity (as measured by the difference in the proportion of match responses for more reliable cue vs. less reliable cue trials) on the last 20 trials as a function of which strategy was reported to have been used last. Only those strategies with *n*s greater than 3 were included. The results were collapsed across conditions to increase the *n*s and because the predicted cue sensitivity should be the same in all conditions. As the RCCL model predicted, participants showed almost no cue sensitivity when their reported strategies were unrelated to the cue color (e.g., the match strategy), whereas they did show cue sensitivity when their reported strategies were related to the cue color.[5] This trend was maintained, although a bit weaker, even when the same analysis was performed separately for the two- and four-color conditions.

[5] It is likely that the weak sensitivity to sample type for the random strategy was due to the participants not using it for all of the last 20 trials.

Thus, the average weak sensitivity to cue in the two-color condition, as both we and Goodie and Fantino (1995) found, was not due to an overall absence of cue sensitivity but to the fact that the two-color task tends to lead participants to adopt task representations that do not include the cue. This finding points to the perils of aggregate analyses that average across strategies (cf. Siegler, 1987).

The results of Experiment 1 also relate to previous results in the base-rate literature. Specifically, the main finding of Experiment 1—that participants were sensitive to base rates on test trials in the BST—seems to contradict the fairly robust finding that in experiential choice tasks people exhibit base-rate sensitivity during training trials but not during test trials (e.g., Estes et al., 1989; Gluck & Bower, 1988). What explains this difference, and why does the RCCL model predict the sensitivity we observed in Experiment 1?

Our explanation rests on the RCCL model's view of base-rate effects in terms of choice among learned strategies. In Experiment 1, to complete the training trials, participants had to construct a set of strategies for choosing the first stick in their solutions. According to RCCL, participants would learn the relative success rates of these strategies (and possibly new ones as well) so that they would eventually prefer those with higher success rates, demonstrating overall base-rate sensitivity during training. When the test trials of Experiment 1 were presented, participants would be faced with novel problems. (As in other research, these problems were designed not to overlap with the training problems.) The test problems in our experiment were presented in terms of the same set of features as the training problems, so the strategies that participants had learned and used during training would apply perfectly well to the test problems. Thus, according to the RCCL model, because participants would be using their learned strategies at training and testing, whatever preferences participants acquired during training should transfer to these test trials, leading to base-rate sensitivity at testing. In contrast, previous experiments showing base-rate neglect at testing have usually presented test trials with a different (typically reduced) set of features relative to the training trials. According to RCCL, because a person's strategies tend to be based on a limited set of features, this difference between training and testing could create a situation in which participants' learned strategies do not apply to the test trials. When this is the case, success-based preferences among the learned strategies will not transfer to test trials because the strategies themselves do not transfer. Thus, RCCL predicts that when the same strategies apply at training and testing, there should be no difference in base-rate effects but that when training strategies do not apply at testing, the strategy preferences learned during training will not transfer.

We can test the first half of this prediction directly by comparing participants' base-rate sensitivity at training and testing in Experiment 1. We compared these base-rate effects at the individual level to avoid averaging out potential differences. Specifically, for each participant, we computed two proportions: the proportion choice of the more successful procedure over the last 20 training problems and the proportion choice of the more successful procedure over the final test problems. Of all 58 participants, only 3 showed a significant difference in these choice proportions. This is approximately the number of statistically significant cases one would expect under the null hypothesis of no difference between training and testing proportions (i.e., 3 of 58 participants; ≈5%).[6] This result of numerically indistinguishable base-rate effects at training and testing conforms to the RCCL model's prediction of no differences. Moreover, given the RCCL explanation, this result does not contradict previous findings of different base-rate effects at training and testing because, according to the RCCL model, similar base-rate effects depend on the applicability of training strategies to test trials.

Task Representations

In Experiments 1 and 2, we argued that task representations are at the heart of learning to use base rates and cue-specific information, and yet we measured only behavioral choice patterns and explicit strategy reports, observable indicators of the underlying task representations. Although differences in representations are consistent with our account, the evidence for representational change is indirect. However, this issue is not unique to our efforts. Similar criticisms could be applied to the pioneering work of Newell and Simon (1972) on task representations in problem solving and Kaplan and Simon's (1990) work on representation change. Because task representations refer to mental states, they remain outside of our grasp to measure directly, barring radical improvements in brain-imaging technology.[7] Thus, we must rely on the assumptions that (a) behavioral patterns and explicit reports that make use of features are evidence of those features being part of a task representation and (b) behavioral patterns and explicit reports that do not make use of features are evidence of those features not being part of a task representation. Yet, it remains possible that features for which there is no outward indication are nonetheless represented by the individual.

The fundamental importance of task representations in problem solving has been understood for quite some time (e.g., Kaplan & Simon, 1990; Kotovsky, Hayes, & Simon, 1985; Newell & Simon, 1972). However, task representations have typically been assumed to remain stable over the course of problem solving. In contrast, we argue that task representations are ever-changeable throughout problem solving. We found evidence for such representation variabil-

[6] We computed statistical significance according to a z test for proportions with $\alpha = .05$. Note that the number of participants with a statistically significant difference between training and test proportions was the same regardless of whether the test proportions were computed based on all 10 test problems or based on the 6 test problems designed to be especially different from the training problems.

[7] Other techniques, such as the study of concurrent verbal protocols, eye movements, and gestures accompanying speech, provide other useful, but still indirect, measures of problem solvers' representations (e.g., Alibali, Bassok, Solomon, Syc, & Goldin-Meadow, in press; Alibali & Goldin-Meadow, 1993; Ericsson & Simon, 1993).

ity in two simple tasks and expect the variability to be at least as great in more complex situations, especially when one allows for negotiation of task representations in group problem-solving situations (e.g., Garrod & Doherty, 1994).

Our conception of task representation is related to Burns and Vollmeyer's (1996) concept of a model space and to Schunn and Klahr's (1995, 1996) concept of a data representation space. Burns and Vollmeyer postulated that, in addition to considering the set of possible problem-solving states and rules for moving from one state to the next, people also consider different models for the structure of the problem-solving task. Burns and Vollmeyer's concept of models is hypothesized to constrain the set of possible problem-solving states as well as the set of rules for moving between states. Similarly, our conception of a task representation is used to constrain problem solving; however, our task representations are defined concretely as a filter on the set of features used to represent problem-solving states and define strategies, whereas their models are less well defined. In contrast, Schunn and Klahr's data representation space is more similar to our task representation, although there the focus is on scientific discovery rather than on problem solving more generally. Their data representations are defined as the set of objects and object features that scientists use to characterize data from experiments. Like task representations, data representations act as a filter on the features used to represent environmental input. Of particular interest to the claims in this article, they also found high levels of variability in the features used by individuals from one trial to the next. Their task was a complex discovery microworld, suggesting that our findings will generalize to more complex tasks.

The fundamental role that task representations play in choice highlights the issue of their origins. Initially, feature perceptual saliency and background knowledge are used to select features. In Experiment 2 we manipulated feature saliency and thereby influenced the participants' task representations and choice behavior. Over time, as Experiment 1 demonstrated, feature predictiveness becomes more important, with unpredictive features being dropped. However, feature saliency remains influential, as the differences in final behavior between the two- and four-color conditions of Experiment 2 demonstrated.

Coming Full Circle: Relating the Textual and Experiential Paradigms

We introduced the problem of choosing in an uncertain world by citing results from the textual paradigm, in which participants are given explicit and separate information on the base rates of different options and the predictiveness of problem-specific cues. In these experiments, contrary to initial findings, there has been surprising variability in the direction of base-rate effects: Some experiments show base-rate neglect and others base-rate sensitivity. For example, base-rate sensitivity is improved when the problem's wording emphasizes either the independence of the base-rate and cue-specific information (Macchi, 1995), the random sampling of cases (Gigerenzer et al., 1988), or the apparent relevance of the base rates (Ajzen, 1977; Bar-Hillel, 1980; Beckett & Park, 1995; Birnbaum & Mellers, 1983; Carroll & Siegler, 1977; Fischhoff et al., 1979). In each case, the manipulations of the problem statement wording lead to certain features being emphasized, which then change how people represent and process the problem. At an abstract level, these effects are similar to the experiential paradigm effects explained by the RCCL model: Which features are included or emphasized in the task representation determines the degree of base-rate sensitivity.

The way in which representation differences influence base-rate sensitivity can, however, vary between the textual and experiential paradigms. For example, Gigerenzer and Hoffrage (1995) found that people show stronger base-rate sensitivity when the problem statements present information in terms of frequencies (e.g., 1 in 1,000) instead of probabilities (e.g., .001). Under Gigerenzer and Hoffrage's analysis, such frequency formats are easier to process because they require fewer and simpler computations. Consider a standard frequency-format version of Tversky and Kahneman's (1982) taxicab problem. Instead of the version presented in our introduction, the problem may be presented as follows: (a) Eighty-five of every 100 taxicabs in the city are green; (b) 12 of every 15 blue taxicabs are seen as blue by a witness; and (c) 17 of every 85 green taxicabs are seen as blue by a witness.

Computing the probability that the taxicab that was seen as blue by the witness really was blue involves calculating the relatively simple ratio of (blue taxicabs seen as blue)/(all taxicabs seen as blue) = 12/(12 + 17). In its original probability format, however, this problem involves applying Bayes's theorem, which involves more complicated computations:

$$P(\text{blue cab}\,|\,\text{blue report}) = \frac{P(\text{blue cab})P(\text{blue report}\,|\,\text{blue cab})}{P(\text{blue cab})P(\text{blue report}\,|\,\text{blue cab}) + P(\text{green cab})P(\text{blue report}\,|\,\text{green cab})}.$$

A similar analysis of computational complexity can be applied to the RCCL model's processing in the experiential paradigm. Note that the experiential version of this problem would include a sequence of trials instead of a problem statement, with each trial providing the witness's report as a cue and asking the participant to choose the taxicab's true color. Here, the desired probability from the textual problem simply corresponds to the success rate of the strategy "if the cue is blue, then choose blue." Because the RCCL model claims that people are always learning strategy success rates through experience, any individual who has been using this strategy would be able to display sensitivity to the different taxicab colors' base rates. No extra computation is necessary. However, the situation is much different for individuals who do not include specific cue and choice colors in their representations but instead use strategies such as "match the cue" and "do not match the cue." Such individuals would not have learned the relevant success rate for this problem and hence would be ill-equipped to display the desired base-rate sensitivity. Instead, these individuals would have learned the success rate of matching versus not matching. Mapping back to the textual paradigm, this corresponds to learning that 80% of the time the witness is reliable and 20% of the time unreliable. Under this match/not-match task representation, there is not even a question about the relative complexity of producing base-rate-sensitive behavior; this representation does not offer enough information to compute the "desired" base rates. As the RCCL model claims, people who do not represent the relevant features of a task will not be able to display base-rate sensitivity to those features. The key insight is that different task representations aggregate experience in ways that can change the informational content available (or not available) for producing base-rate sensitivity.

Thus, our work on base-rate effects in the experiential paradigm essentially leads to the same conclusion as the research on textual base-rate effects: Representation of the problem information is a critical factor in determining base-rate sensitivity versus neglect. The effects of representation in the two paradigms, however, are different. In the textual paradigm, representational format affects how a person processes the base-rate and problem-specific information. In the experiential paradigm, we have argued that task representation—the set of features included in one's mental representation—affects which base rates a person can learn from experience and hence the base-rate effects exhibited in choice behavior.

This is not the only relationship between the two paradigms. It is also the case that moving beyond aggregate analyses to "individualized" analyses has shed light on base-rate effects in the textual paradigm. For example, Stanovich and West (1998) studied individual differences in base-rate neglect and other reasoning fallacies observed on text-based problems. They found that cognitive ability measures explained some of the variance of people's performance on these tasks. In addition, Gigerenzer et al. (1988) found that computing base-rate effects at the individual level substantially reduced the degree of measured base-rate neglect. That is, taking into account an individual's probability judgments across different problems corrects for individual differences and focuses on the individual's true sensitivity to different base rates. In the experiential results reported in this article, we have found similar advantages to studying choice at the individual level. In particular, we compared training and test base-rate effects at the individual level and found them to be indistinguishable when the same representation and strategies could be used. We also studied the amount of change in choice within individuals. These analyses highlight the importance of individual differences in various stages of the RCCL model. More generally, the themes of representation and variability that are now common to the textual and experiential paradigms testify to the importance of these issues in furthering the understanding of human choice processes.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35,* 303–313.

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (in press). Illuminating mental representations through speech and gesture. *Psychological Science.*

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about the child's state of mind. *Cognitive Psychology, 25,* 468–523.

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

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233.

Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin, 21,* 21–31.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verablizable knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36A,* 209–231.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who can vary in credibility. *Journal of Personality and Social Psychology, 45,* 792–803.

Burns, B. D., & Vollmeyer, R. (1996). Goals and problem solving: Learning as search of three spaces. In G. W. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 23–26). Mahwah, NJ: Erlbaum.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General, 121,* 177–194.

Carroll, J. S., & Siegler, R. S. (1977). Strategies for the use of base-rate information. *Organizational Behavior and Human Performance, 19,* 392–402.

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

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as Data* (Rev. ed.). Cambridge, MA: MIT Press.

Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). New York: Academic Press.

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 556–571.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance, 23,* 339–359.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. (1964). Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 250–316). Stanford, CA: Stanford University Press.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53,* 181–215.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513–525.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instructions: Frequency formats. *Psychological Review, 102,* 684–704.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52,* 464–473.

Gluck, M. A., & Bower, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247.

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition, 25,* 237–255.

Goodie, A. S., & Fantino, E. (1995). An experientially derived base-rate error in humans. *Psychological Science, 6,* 101–106.

Goodie, A. S., & Fantino, E. (1996). Learning to commit or avoid the base-rate error. *Nature, 380,* 247–249.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22,* 374–419.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–53.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology, 17,* 248–294.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8,* 225–247.

Lovett, M. C., & Anderson, J. R. (1995). Making heads or tails out of selecting problem-solving strategies. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 265–270). Hillsdale, NJ: Erlbaum.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology, 31,* 168–217.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40,* 287–298.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A,* 188–207.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 288–301.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology, 38,* 231–248.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

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

Schunn, C. D., & Klahr, D. (1995). A 4-space model of scientific discovery. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 106–111). Hillsdale, NJ: Erlbaum.

Schunn, C. D., & Klahr, D. (1996). The problem of problem spaces: When and how to go beyond a 2-space model of scientific discovery. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 25–26), Mahwah, NJ: Erlbaum.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250–264.

Spellman, B. A. (1996). Implicit use of base rates in experiential and ecologically valid tasks (Commentary on J. J. Koehler, The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges). *Behavioral and Brain Sciences, 19,* 38.

Stanovich, K. E., & West, R. F. (1998). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 26,* 161–179.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). New York: Cambridge University Press.

Received May 12, 1997
Revision received October 7, 1997
Accepted March 27, 1998 ■