Research Report

Modeling Experimentally Induced Strategy Shifts

Scott Brown,[1] Mark Steyvers,[2] and Pernille Hemmer[2]

[1]School of Psychology, University of Newcastle, Callaghan, New South Wales, Australia, and [2]Department of Cognitive Science, University of California, Irvine

ABSTRACT—*In dynamic decision-making environments, observers must continuously adjust their decision-making strategies. Previous research has focused on internal fluctuations in decision mechanisms, without regard to how these changes are induced by environmental changes. We developed a simple paradigm in which we manipulated task difficulty, thereby inducing changes in decision processes. We applied this paradigm to recognition memory, manipulating task difficulty by changing the similarity of lures to targets. More difficult decision environments caused participants to make more careful decisions, but these changes did not appear immediately. We propose a simple theoretical account for these data, using a dynamic version of signal detection theory fitted to individual subjects. Our model represents a significant departure from existing models because it incorporates subject-controlled parameters that may adjust over time in response to environmental changes.*

What triggers a participant to be more careful in one condition than another, or to pay more attention to some features than others? These questions are important in considering real-world environments, where changes are frequent and require constant adjustment of decision-making strategies. Most laboratory analyses of such changes are static: The conditions are considered separately, with the aim of establishing differences in decision-making strategies regardless of how environmental changes give rise to those differences. In this article, we illustrate a method for going beyond this level, by using dynamic analyses.

Our approach is substantially different from related work in dynamic systems (e.g., Gilden, Thornton, & Mallon, 1995; Kelly, Heath, & Longstaff, 2001; Van Orden, Holden, & Turvey, 2003). That research has focused on spontaneous rather than experimentally induced changes. For example, a common paradigm involves collecting a long, stationary sequence of data and extracting information about autocorrelation properties, or low-dimensional nonlinear systems that might have produced the data. These quasi-experimental analyses are post hoc—one cannot predict the location of changes in the data without first looking at those very data. A fundamentally different approach is to *induce* dynamic changes via experimental manipulations. This approach is a priori; changes in data can be predicted to occur around the points at which experimental conditions were changed.

We investigated induced dynamic effects in recognition memory. We presented subjects with a series of photographic images of everyday objects to be studied. Later, we tested the subjects with a mixture of images presented at study (targets) and new images (lures). The crucial experimental manipulation was an increase (or decrease) in difficulty halfway through each test phase, caused by increased (or decreased) similarity of targets and lures. This manipulation forced subjects to be more (or less) careful, just as one must be more careful when recognizing a child's face in a crowd of children (and less careful when recognizing a child's face in a crowd of adults). Our primary research questions were, can we measure the onset of corresponding changes in subjects' strategies, and if so, how quickly are those changes made? We developed a dynamic version of signal detection theory (SDT) to explore these questions.

METHOD

We created two different levels of difficulty by changing the kind of distractors (lures). In the easy condition, lures were quite clearly different from the studied items; in the difficult condition, lures were very similar to the studied items. Our experimental design was similar to that of Benjamin and Bawa's (2004) Experiment 3, with two important differences. First, we adjusted our test conditions to increase difficulty (e.g., average of around 63% correct responses, compared with around 90%

Address correspondence to Scott Brown, School of Psychology, Psychology Building, University of Newcastle, Callaghan, NSW 2308, Australia, e-mail: scott.brown@newcastle.edu.au.

for Benjamin & Bawa's study). Second, we had each participant change between the easy condition and the difficult condition six times, whereas Benjamin and Bawa's participants experienced just one change. Our design allowed us to make more reliable measurements of performance around the point of stimulus change and thus to analyze dynamic-model hypotheses.

We created the memory stimuli from images of everyday objects photographed against white backgrounds. There were 30 images in each of six categories: telephones, bags, hammers, cameras, binoculars, and guitars (see Fig. 1 for an example).

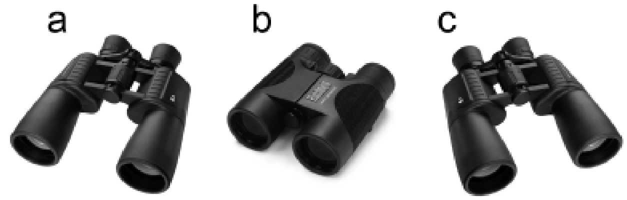

Fig. 1. Sample study picture (a) with sample distractors from the easy condition (b) and the difficult condition (c).

Images in a study list were all drawn from a single category. During the test phase, lures were either difficult or easy. Easy lures were simply new pictures drawn from the same category used in the study list. For example, Figure 1a shows a picture of a pair of binoculars that could have been used as a studied item. Figure 1b shows a picture of a different pair of binoculars, not shown in the study list—a lure for the easy condition. Difficult lures were mirror images of studied pictures: For example, Figure 1c shows the difficult lure corresponding to Figure 1a. We never tested both mirror and nonmirror versions of any image. We warned subjects that mirror images of studied items might appear as test items, and instructed subjects that mirror images were not old items and should be responded to as lures. Each study list was created by drawing 18 images from a single category. The corresponding test list was created by drawing 12 images from the study list (old items), 6 new images from the same category (easy lures), and 6 mirror images of studied items (difficult lures). The remaining 6 items in the category were used for 3 initial and 3 final buffer pictures on the study list, but were not used in the test list. Figure 2 has an illustrative example of a study and a test list, showing a transition from the easy context (the first 12 test trials) to the difficult context (the final 12 test trials).

During the test phase, subjects were shown pictures (either studied items or lures) one at a time and asked to rate how confident they were that each picture had been previously presented in the study list. They responded using a mouse and an on-screen rating scale. The rating scale consisted of three choices for both old and new: "very sure," "sure," and "not sure"; in all analyses reported here, we collapsed these ratings into "old" and "new" responses. Participants were not given any feedback about the accuracy of their decisions. Each participant completed six study-test blocks. During each block, the difficulty condition switched after the 12th trial, such that there were only easy lures for Trials 1 through 12 and only difficult lures for Trials 13 through 24, or vice versa.

Blocks containing easy-to-difficult and difficult-to-easy switches alternated, with the direction of change for the first block selected randomly. The 13th image in each test list was a lure, so that the switch point could be defined accurately. Participants were explicitly warned to be vigilant against classifying mirror images of studied items as old. Participants were not told that easy and difficult lures would occur in different sections of the experiment, or that there would be switches between easy and difficult conditions. In fact, postexperimental debriefings suggested that participants were unaware of the separation of easy and difficult lures into different parts of each block. We also did not explicitly tell subjects that each test list would contain exactly 50% old and 50% new items. However, we assume that our subjects expected this property to hold, and may have used this knowledge during the experiment.

Forty-seven undergraduates from the University of California, Irvine, participated in the study, but data from 11 subjects were omitted from analysis. Those 11 participants did not pay attention to the task instructions and treated mirror-image lures as old items, yielding unacceptable false alarm rates (FARs; over 75%).

RESULTS

The bottom left panel of Figure 3 shows changes in the hit rate (HR) and the FAR across the 24 trials within each block, separately for blocks that began with the easy condition and ended with the difficult condition and blocks that began with the difficult condition and ended with the easy condition. In the first 6

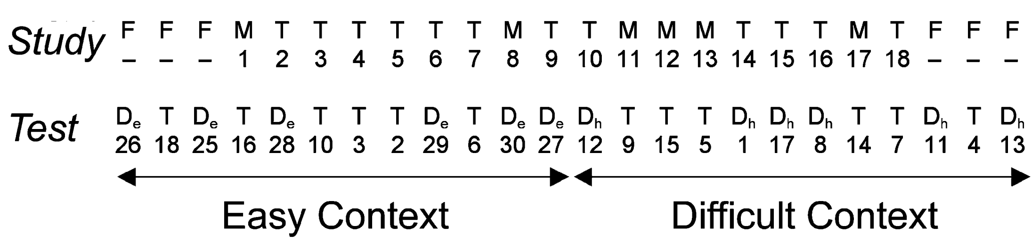

Fig. 2. Illustration of the sequence of stimuli in the study and test blocks. The numbers below the items indicate the image numbers. F = filler item; T = study item; M = study item whose mirror image was to be used as a hard distractor at test; D_e = easy distractor; D_h = hard distractor.

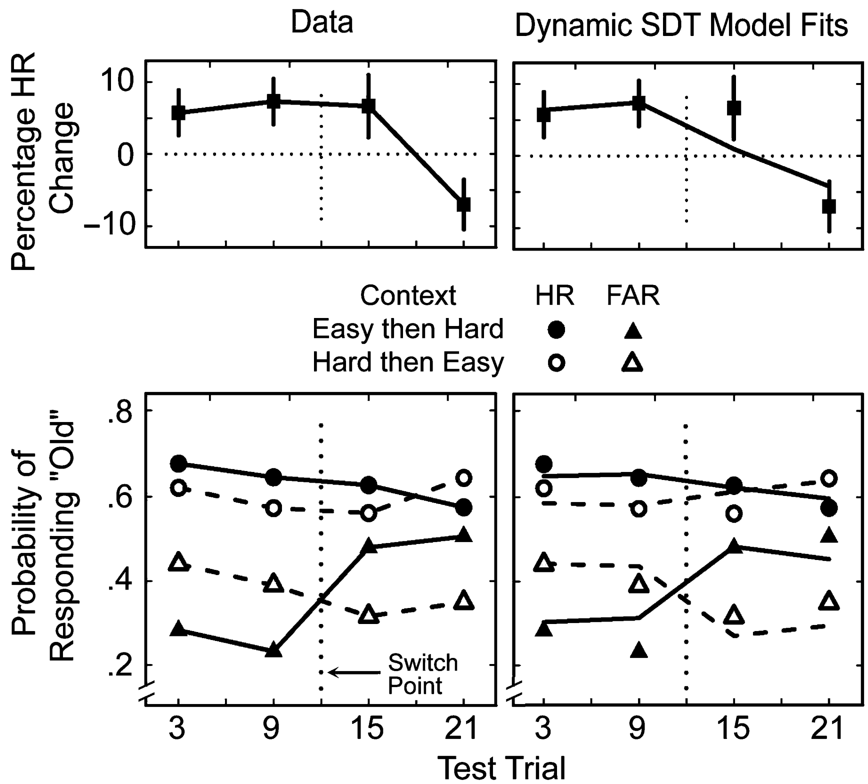

Fig. 3. Experimental data (left) and fits of the dynamic signal detection theory (SDT) model to those data (right). The lower graphs show mean hit rates (HR) and false alarm rates (FAR) across trials within each block, separately for blocks that started in the easy context and switched to the difficult context and for blocks that started in the difficult context and switched to the easy context. The upper panels show the differences in HR between the easy and difficult contexts (calculated by subtracting HR for blocks that began with the difficult context from HR for blocks that began with the easy context). The error bars in the upper panels show standard errors, calculated using a repeated measures analysis of variance algorithm. In the panels showing the model-fitting results, the lines show predictions from the dynamic SDT model, calculated separately for each subject and then averaged (the data points are from the graphs on the left). For these graphs, we averaged data across participants and within six-trial windows, but for the model analyses reported in the text, we used unaveraged data.

trials of each block, the FAR was higher and the HR was lower in the difficult context than the easy context. This is the well-known mirror effect, which has been very important in the development of memory theories (e.g., Benjamin & Bawa, 2004; Glanzer & Adams, 1990; Glanzer, Adams, Iverson, & Kim, 1993; Stretch & Wixted, 1998). A higher FAR was to be expected in the difficult condition because the difficult lures were very similar to the studied items, so many of them were erroneously called "old." The upper left panel of Figure 3 plots the difference between the HRs in the easy and difficult conditions. The fact that the HR was lower in the difficult condition than in the easy condition before the change point must be ascribed to internal differences within the participants, as the studied items producing the HRs were the same in the easy and the difficult contexts. Most researchers assume that some kind of criterion shift underlies the mirror effect (at least, when the criterion is not required to shift on every trial—see Stretch & Wixted, 1998). That is, when a decision environment becomes more difficult, or the probabilities of the two stimulus classes change, observers respond by changing the amount of evidence they require to make each response type (e.g., Benjamin & Bawa, 2004; Berch, 1976; Marken & Sandusky, 1974).

After the stimulus switch point (between Trials 12 and 13), the FAR changed dramatically, reflecting the change in lures. The FAR for transitions from easy to difficult lures showed a dramatic increase, and the FAR for difficult-to-easy transitions also showed a significant decrease. These changes occurred immediately after the switch point, but there was no correspondingly immediate change in the HR. After the switch point, the HR for blocks that began in the difficult condition remained lower than the HR for blocks that began in the easy condition, even though those conditions had reversed. Thus, immediately after the change, the HR was higher in the difficult condition than in the easy condition—the opposite of a mirror effect. Finally, during the last six trials of each block, the ordering of the HRs reversed, and the mirror effect re-appeared.

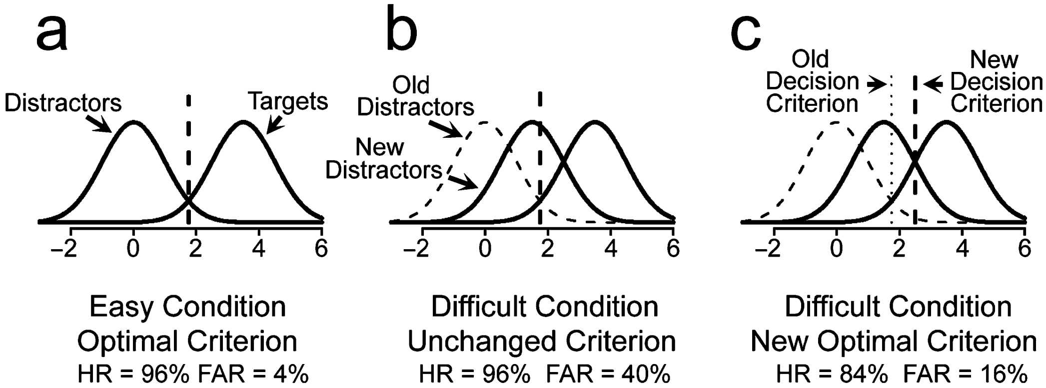

Fig. 4. Illustration of the dynamic signal detection theory (SDT) model during a change from the easy to the difficult condition. The example begins with a standard SDT model for an easy distinction (a); the x-axis represents familiarity, and probability distributions for targets and distractors are shown, along with the decision criterion for responding "old" versus "new" (the dashed vertical line). When the distractors become more similar to the targets, the decision becomes more difficult, but the criterion does not change immediately (b). As a result, the false alarm rate (FAR) increases, but the hit rate (HR) is unchanged. After some lag time (c), the criterion is updated, decreasing both the HR and the FAR.

DYNAMIC SDT

Our results show that after the onset of an experimental change in the difficulty of distractors, the usual mirror pattern in HR and FAR took 6 to 12 trials to reestablish. The parallel changes for HR and FAR suggest a criterion-shift explanation for this mirror effect. To provide such an explanation, we have developed a simple SDT model that includes a dynamic lag component to model the delay in reestablishing mirror effects. In a previous study involving a different paradigm (Brown & Steyvers, 2005), we used this model to estimate parameters describing performance, including sensitivity (d'), bias, and change point. The dynamic model we employ here incorporates simple SDT models to describe performance in the easy and difficult conditions. The model is dynamic in the way it accounts for changes between the two conditions: We assume that when the lures change, the lure distribution in the model changes immediately, but that the decision criterion changes only after some time lag (specified by the change-point parameter). This process is illustrated in Figure 4.

Figure 4a shows a standard SDT model for the easy condition. Note that the distractor distribution is well separated from the target distribution, because the distractors are easy. In Figure 4b, the distractors are much more difficult, so the distractor distribution is much closer to the target distribution. This change increases the FAR immediately and dramatically. Because the decision criterion has not yet moved from its previous placement (which was appropriate for the easy condition), the HR remains unchanged. After some time has passed (a lag), the decision criterion is updated, as shown in Figure 4c. This change causes parallel decreases in the FAR and HR. The key dynamic assumption is that the properties of the target and lure distributions are fixed by the experimenter, but the location of the decision criterion is under subjects' control.

In Figure 4, we simplified the model description by using decision criteria that were set optimally. However, when we fit the model to the data, we freely estimated the decision criteria (bias parameters).[1] The predicted HR and FAR averaged across subjects are shown in the bottom right panel of Figure 3, together with the averaged data. The model fit the data quite well, capturing the large, immediate change in the FAR after the stimulus switch point, as well as the slower change in the HR (shown in detail in the upper right panel of Fig. 3). The dynamic SDT model also captured the slower postswitch trends in the FAR, that is, the pattern of slow change in the direction opposite the direction of the initial large changes.[2] Reasonable estimates of

[1]We used maximum likelihood to estimate five model parameters for each subject, without averaging over trials or blocks. That is, for given parameters, we calculated expected HR and FAR from the model, then used these values to calculate the likelihood of 1 subject's data sequence under the model. We repeated this process, adjusting parameter estimates using the simplex algorithm, to maximize the likelihood. Note that in the simple static case, this method is equivalent to calculating sensitivity and bias using the standard z transforms.

[2]The model assumes that the variances of the distributions for old and new items are equal. Previous research suggests that this assumption is often untrue, especially in recognition memory (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Malmberg, 2002; Ratcliff, Sheu, & Gronlund, 1992; Sheu & Heathcote, 2002; Verde & Rotello, 2003). Because our model includes a criterion shift during otherwise-stationary data sequences, we were able to directly estimate variance ratios. This approach uses the same principle as receiver-operating-characteristics (ROC) analysis, except that it involves two explicitly different criterion placements, rather than several implicit criteria used to model confidence judgments. For most participants, the unequal-variance dynamic SDT model did not fit the data significantly better than the simpler equal-variance model (according to a nested-model comparison using a chi-square test on log-likelihood differences). The mean estimated ratio of the standard deviation for new items to the standard deviation for old items was 0.53 (standard error = 0.31). Note that this estimate of the variance ratio was obtained without using the confidence-rating data, and yet it is approximately consistent with estimates based on ROC analyses of those data. The agreement between our dynamic estimates and ROC estimates suggests that our (new) methodology for estimating variance ratios in SDT produces reasonable values; however, either the data or the effect sizes were insufficient to make reliable distinctions between equal- and unequal-variance models.

lags were obtained: It was estimated that, on average, participants changed their criterion location after the 15th trial of each block, 3 trials later than optimal. That the lag estimates are reasonable, and the fits to the data are good, suggests that the model captures behavior well even when fit to relatively scant data. However, more observations of switch points would improve parameter estimation (e.g., using a simpler task, we collected either 10 or 20 observations per subject—Brown & Steyvers, 2005).

DISCUSSION

We created a recognition memory task that switched between easy and difficult conditions by changing the properties of lures in the middle of each test list. Results were similar to those of other studies (Benjamin & Bawa, 2004; Stretch & Wixted, 1998), as we observed a mirror effect consistent with changes in decision criteria. We applied to the data a dynamic SDT model in which criteria shifted in response to changes in the properties of lures, but these shifts occurred some time later than the stimulus changes.

The dynamic SDT model is our attempt to illustrate a weakness in many current models. In most models, some parameters—such as decision criteria—are assumed to be under subjects' control, and these parameters are most often simply assigned some value, without any explanation of when that value changes after conditions are changed. The basic principle behind the dynamic SDT model (an assumption that can be implemented in many other models as well) is that different decision-making strategies are adopted in response to changes in the decision-making environment, but that these changes in behavior occur at some lag, which can be estimated from the data. These estimated lags can be of both psychological and methodological interest.

Our use of dynamic analyses for dynamic designs is unusual, but not unprecedented. We concentrated on cognitive changes that take place over many (6–12) trials. Other researchers have examined shorter-term dynamics, most notably in studies of sequential effects during decision making or categorization (e.g., Petrov & Anderson, 2005; Stewart, Brown, & Chater, 2005; Treisman & Williams, 1984; Vickers & Lee, 1998, 2000; Ward & Lockhead, 1970). These studies differed from our work in two important ways. First, the sequential effects were much shorter-term fluctuations than the changes we studied—on the order of 1 or 2 trials, rather than 6 to 12. Second, the sequences of stimuli assumed to cause sequential effects were always confounded with sequences of responses, making causal inferences difficult. The design we have used avoids this problem, and is not limited to recognition memory.

Acknowledgments—This work was supported by a grant from the U.S. Air Force Office of Scientific Research (Grant FA9550-04-1-0317).

REFERENCES

Benjamin, A.S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*, 159–172.

Berch, D.B. (1976). Criterion change in continuous recognition memory: A sequential effect. *Bulletin of the Psychonomic Society*, *7*, 309–312.

Brown, S.D., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 587–599.

Gilden, D.L., Thornton, T., & Mallon, M.W. (1995). 1/f noise in human cognition. *Science*, *267*, 1837–1839.

Glanzer, M., & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5–16.

Glanzer, M., Adams, J.K., Iverson, G.J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546–567.

Glanzer, M., Kim, K., Hilford, A., & Adams, J.K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 500–513.

Kelly, A., Heath, R.A., & Longstaff, M. (2001). Response-time dynamics: Evidence for linear and low-dimensional nonlinear structure in human choice sequence. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology A*, *55*, 805–840.

Malmberg, K.J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 380–387.

Marken, R.S., & Sandusky, A.J. (1974). Stimulus probability and sequential effect in recognition memory. *Bulletin of the Psychonomic Society*, *4*, 49–51.

Petrov, A., & Anderson, J.R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*, 383–416.

Ratcliff, R., Sheu, C.-F., & Gronlund, S.D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518–535.

Sheu, C.-F., & Heathcote, A. (2002). A nonlinear regression approach to estimating signal detection models for rating data. *Behavior Research Methods, Instruments, & Computers*, *33*, 108–114.

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

Stretch, V., & Wixted, J.T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1379–1396.

Treisman, M., & Williams, T.C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68–111.

Van Orden, G.C., Holden, J.G., & Turvey, M.T. (2003). Self organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*, 331–350.

Verde, M.F., & Rotello, C.M. (2003). Does familiarity change in the revelation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 739–746.

Vickers, D., & Lee, M.D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module.

Nonlinear Dynamics, Psychology, and Life Sciences, *2*(2), 169–194.

Vickers, D., & Lee, M.D. (2000). Dynamic models of simple judgments: II. Properties of a self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multichoice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences*, *4*(3), 1–31.

Ward, L.M., & Lockhead, G.R. (1970). Sequential effect and memory in category judgment. *Journal of Experimental Psychology*, *84*, 27–34.

(RECEIVED 11/21/05; REVISION ACCEPTED 6/22/06;
FINAL MATERIALS RECEIVED 6/30/06)