Learning to Adapt Evidence Thresholds in Decision Making

Ben R. Newell (ben.newell@ unsw.edu.au)
School of Psychology,
University of New South Wales,
Sydney, Australia

Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences,
University of California, Irvine,
Irvine, CA, 92697-5100, USA

Abstract

A basic challenge in decision-making is to know how long to search for information, and how to adapt search processes as performance, goals, and the nature of the task environment vary. We consider human performance on two experiments involving a sequence of simple multiple-cue decision-making trials, which allow search to be measured, and provide feedback on decision accuracy. In both experiments, the nature of the trials changes, unannounced, several times. Initially minimal search is required, then more extensive search is required, and finally only minimal search is again required to achieve decision accuracy. We find that people, considered both on aggregate, and as individuals, are sensitive to all of these changes. We discuss the theoretical implications of these findings for modeling search and decision-making, and emphasize that they show adaptation to an external error signal must be accompanied by some sort of internal self-regulation in any satisfactory account of people's behavior.

Keywords: evidence accumulation, heuristics, decision-making, learning, self-regulation

Introduction

A problem faced commonly by decision makers is determining how much information to incorporate into a decision. Some decisions are trivial (e.g., choosing a breakfast cereal), but some more important (e.g., choosing a mate), and consequently the amount of information or evidence examined prior to deciding will vary. One way to model this variance is to suggest that people sample evidence *sequentially* and adjust the amount of evidence they consider according to a *decision threshold*. Inherent in this conception is that thresholds will vary not just between decisions but also between individuals (Lee & Cummins, 2004; Vickers, 1979). Newell (2005) suggested an 'adjustable spanner' (or wrench) to capture this idea; a spanner in which the width of the jaws represents the amount of evidence a person accumulates before making a decision. In this paper we develop this perspective by examining how people learn to adjust their evidence thresholds in dynamic decision environments.

Adapting to changing environments

Some recent studies examining adaptation to changes in the statistical structure of decision environments have led to rather pessimistic conclusions. Bröder and Schiffer (2006) considered environments in which either a *compensatory* strategy (one which weights and integrates all cue information) or a *noncompensatory* strategy (one which considers only a subset of the information) was optimal with respect to the expected monetary pay-off. Participants detected the appropriate strategy in whichever environment they encountered initially, but when the environment changed (mid-way through the experiment) most participants retained the same strategy despite its sub-optimality in the new environment. Bröder and Schiffer explained these results in terms of the application of maladaptive routines: participants used a 'top-down' deliberative mode of thinking in the initial phase to work out the appropriate strategy, but then when this appeared to work successfully slipped into a more 'bottom-up' routine mode and thus failed to test the consequences of applying the strategy on each trial when the environment changed.

In a similar vein, Rieskamp (2006) used decision environments in which the lexicographic strategy Take-the-Best (TTB; Gigerenzer & Goldstein, 1996) was more or less adaptive (in terms of expected monetary pay-off) than a weighted additive strategy (WADD; e.g., Payne, Bettman & Johnson, 1993). In the noncompensatory environment TTB led to 83% correct predictions compared to WADD's 60%; and these values were reversed in the compensatory environment. Participants who transitioned from the non-compensatory environment to the compensatory one showed a distinct *inertia*, resisting the change to the more adaptive strategy. Those transitioning in the opposite direction showed more of a change (the fit of TTB increased relative to WADD in the final block of trials) but this was tempered by a less pronounced adoption of WADD in the first half of the experiment. The inertia effect was predicted and explained by the Strategy Selection Learning model of Rieskamp and Otto (2006). The model states that adaptation will be very slow because a strategy that is successful in the initial environment will accrue considerable reinforcement, and this reinforcement will only gradually diminish in the novel environment on the rare occasions when a participant explores the potential of the competing strategy.

Taken together, these results suggest that when the statistical structure of an environment changes participants show resistance to adapting the strategies they employ (or the level of evidence considered) for making their decisions.

Learning to adapt: exploitation vs. exploration

One of the great discoveries of modern cognitive and biological psychology is that learning is driven by the process of error correction or gradient descent. Credit for this insight is usually given to Widrow and Hoff (1960) although it was discovered independently by researchers in several fields (see Newell, Lagnado, & Shanks, 2007, Chapter 11). The basic notion is that learning occurs via the process of trying to minimize the error between an actual outcome and the predicted outcome of a learning episode (e.g., Rescorla & Wagner, 1972; Young & Wasserman, 2005). This assumption of 'supervised learning' is built into many models of learning in decision problems (Yechiam & Busemeyer, 2005). However, error-correction is not the only way in which organisms learn about their environment. Reinforcement learning contrasts with strictly supervised learning by balancing the exploitation of error-minimization with the exploration of behaviors that can improve the current 'state' of the organism (Young & Wasserman, 2005).

In the context of the studies described above, this exploitation-exploration balance is analogous to the trade-off between accuracy and the effort expended in making a decision. When accuracy is the sole concern, the amount of effort expended is not a factor in determining behavior; however if there is a pressure of cost, or time, or cognitive resources then effort is also considered. The question of exactly how the cost and benefits of accuracy and effort are traded-off against one another has been subject to considerable research, but there is still no consensus on how people learn to adapt their strategy or information acquisition to the environment (Beach & Mitchell, 1978; Bröder & Newell, 2008; Payne et al., 1993; Rieskamp, 2006; Rieskamp & Otto, 2006).

In the studies of Rieskamp (2006) and Bröder and Schiffer (2006) participants were, arguably, able to rely on both error-correction and reinforcement mechanisms to facilitate the transition to more optimal strategies when environmental conditions changed. Error-correction learning was feasible because participants were provided with corrective feedback on each trial, and crucially, the *accuracy* of the different strategies *changed* in the two environments. This means that the 'teaching signal' necessary for supervised learning to occur was present (Young & Wasserman, 2005). Reinforcement learning could have occurred if participants had been willing to engage in sufficient exploration of the environment to discover that an alternative strategy was optimal. The failure to engage in this exploration was explained by Bröder and Schiffer in terms of routinization effects and in terms of 'over-learning', or a too-high expectancy of a strategy's success in Rieskamp's (2006) SSL model.

Can people learn to adapt a threshold?

Our aim was to examine the relative roles of error-correction and reinforcement learning in decision making. We designed a situation in which there were *two* changes in the statistical structure of the environment during the course of the experiment. In an initial block, participants learned in an environment in which the predictions of a noncompensatory strategy (TTB) and a compensatory strategy RAT (for *'rational'*) were identical. Under the assumption that correct inferences based on less information provide greater reinforcement than correct inferences based on more information, one predicts that information search will be reduced when the accuracy of strategies is equated (Rieskamp & Otto, 2006). This should lead to the adoption of TTB-like behavior, or lower evidence thresholds. In the second block, the environment changed so that the RAT strategy now led to more correct predictions than TTB. Thus participants could rely on error-correction learning to adapt their thresholds upwards. In concrete terms, if participants persisted with a low-threshold in block 2 this would lead to a high number of *incorrect* responses. These responses should act as a signal to participants to change their behavior. In block 3 the environment changed again, back to one in which RAT and TTB made identical predictions. Of crucial interest here was whether participants would accumulate less evidence, or whether they would continue with a higher threshold. Note that because the *accuracy* of both strategies was identical in block 3 there is no 'teaching signal' to indicate that a higher threshold is no longer necessary. Thus, if a participant continues to use a high threshold in block 3 she will maintain the same level of accuracy as she experienced by the end of block 2. In order to learn to adapt the threshold in block 3 a participant must engage in some exploration of alternative levels of evidence.

Previous research examining behavior in dynamic environments has tended to focus only on situations in which the *accuracy* of strategies change and in which optimality of a strategy is measured in terms of the expected monetary pay-off. Our study differs from these in that our environment has an initial change signaled by accuracy, but then a second change which can only be learned via exploration and subsequent reinforcement of successful behaviors. In addition, participants did not earn money for correct predictions in our experiment. They were motivated to score highly (the best performing participant was awarded with $15) but the principal motivator was time. In Experiment 1 there was the simple time cost for obtaining information about each cue and in Experiment 2 this cost was exacerbated by introducing a time delay between accessing the cue and being provided with the cue value. Thus 'optimality' was defined in terms

of the opportunity costs related to the time taken to obtain evidence.

In summary, our aim was to examine the role of error-correction and reinforcement learning in an environment with time-costs. Prior research suggests caution in predicting changes in behavior as a result of changes in the environment. We did not provide any indication to participants that the environment would change, nor did any surface features of the experimental task change across the blocks. Thus, any observed threshold changes can only arise from participants' balance between exploitation and exploration of the environment (cf. Bröder & Schiffer, 2006).

Method

Participants

Fifty-nine undergraduate students (Experiment 1 N=30; Experiment 2, N=29) from the University of New South Wales participated in return for course credit.

Stimuli

The experimental environment was created by selecting pairs of objects from the German cities environment used by Gigerenzer and Goldstein (1996). Each object was described by nine binary cues and had an associated criterion value. The cues and criterion were re-described according to a cover story about the search for an energy-efficient fuel source, as described in the Procedure section below.

The sequence of trials was designed in terms of 3 consecutive blocks 50, 100 and 50 trials. For blocks 1 and 3 TTB and RAT made an identical number of correct predictions. In block 2 this was only the case for 50% of trials; on the remaining 50% TTB and RAT made opposite predictions with RAT making the correct prediction in each case, making it the more successful (accurate) strategy. Participants were given no indication of the block structure used to design the trial sequence.

Procedure and Design

The experimental task involved making decisions about which of two objects had a higher criterion value for 200 trials. The task was framed as a search for an energy efficient fuel source. On each trial participants were presented with two samples (A and B) and a selection of nine tests which they could 'run' in order to investigate the samples; the tests included "Does the sample contain Actinium?", "Was the seismic analysis positive?" Clicking on a "RUN TEST" button revealed the answer to each question as either YES or NO.

Each test had a 'success rate' which was a veridical indication of the validity of each test as predictor of whether the sample was richer in the new energy efficient fuel. The success rate for each test was presented on screen and was described to participants as follows: "*if a test has a success of 75% this means that if there were 100 trials in which one sample had a positive result (YES) for that test and the other sample had a negative result (NO) for that test, then the sample with the positive result would be the correct choice (be richer in the energy source) on 75 of those 100 cases, whereas for the remaining 25 cases the other sample would have been richer in the energy source*" (cf. Rieskamp & Otto, 2006). The success rates for the nine tests were: 99, 91, 87, 78, 77, 75, 71, 56 and 51%, as per the cue validities in the German cities environment.

Participants had to run at least one test per trial but were free to choose as many as they liked after that, before making their decision. Following each decision feedback was provided, and a record of how many correct decisions had been made was shown on the screen throughout the experiment. The only difference between Experiment 1 and 2 was that in Experiment 2 there was a time cost to running each test. Specifically, participants had to wait for 3 seconds for the result of each test to be displayed on the screen. During this time a message with the words "Computer now running test" appeared on the screen.

Results

Figure 1 shows the results for one participant in Experiment 1, and is presented to help make clear the structure of the experimental design, and the focus of our analysis. The solid line in Figure 1 shows the pattern of change, expressed as a running weighted average over a small window of trials, in the proportion of extra cues searched. This measure is described in detail below, but basically measures the extent of search on a normalized scale, where TTB-consistent search corresponds to the value 0, and RAT-consistent search corresponds to the value 1. The gray dividing lines show the conceptual division of the trial sequence into three blocks, with blocks 1 and 3 having trials where RAT and TTB make the same predictions, but block 2 having trials where RAT outperforms TTB. Those trials on which the participant made a decision error are shown by crosses.

As Figure 1 shows, the sample participant started by using many tests (i.e., searching many cues), but quickly adapted to search fewer as block 1 progressed. After making a single error after the change of block at trial 50, she began running more tests, to a level consistent with the RAT approach. After trial 150, however, she seems to again reduce her search slightly but consistently, and use fewer tests. Importantly, she did this without having made any errors in the trials around the change from block 2 to block 3. It is these patterns of change in search behavior across the three blocks, at both the group and individual participant level, which are the focus of our analysis.

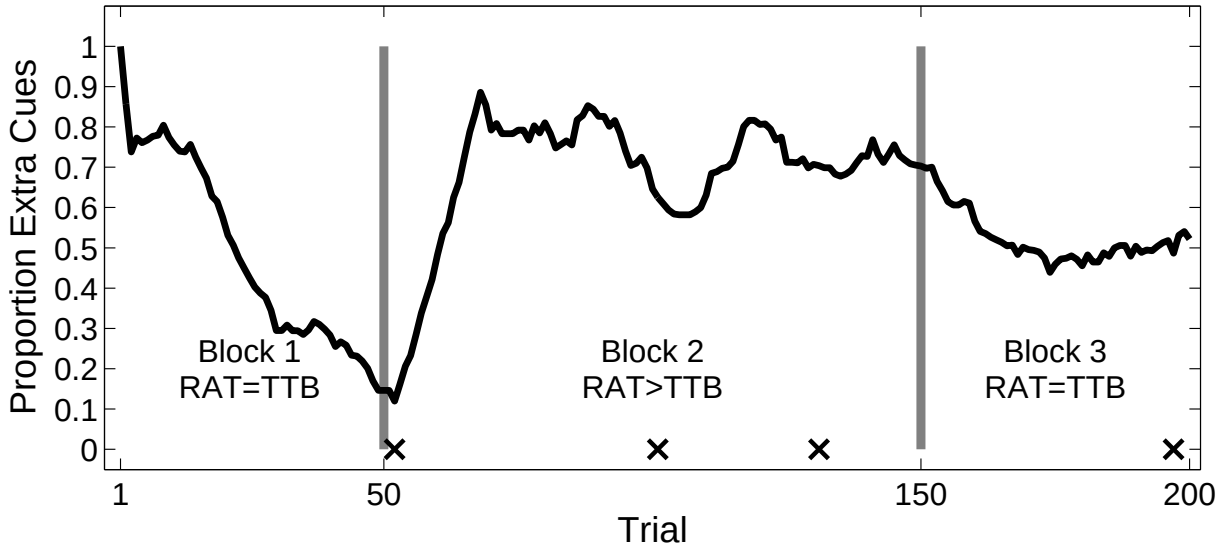

Figure 1: Performance of a sample participant from Experiment 1, showing a weighted running average of the proportion of extra cues searched over the 200 trials. The 3 blocks, differing in the predictive performance of the RAT and TTB heuristics, is shown by the gray lines. Those four trials on which the participant made an error are indicated by crosses.

First, however, Figure 2 displays the accuracy achieved in the three blocks of both experiments. The figure indicates that participants were highly accurate throughout, but experienced a slight decrease in accuracy in block 2 when the environment changed to one in which RAT was the better performing strategy. This pattern was revealed by a quadratic trend for Block; the trend was confirmed by a repeated measures ANOVA which significant in Experiment 2 F (1, 28) = 12.10, p = .002 but not in Experiment 1 F (1, 29) = 2.19, p = .149.

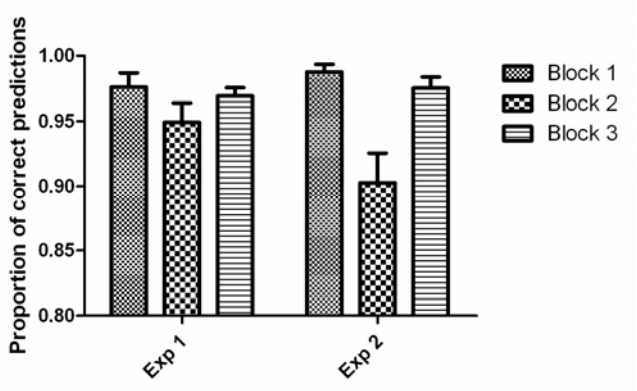

Figure 2: The proportion of correct predictions in each block of Experiments 1 and 2.

Figure 3 shows that the slight decrease in accuracy in block 2 was accompanied by an increase in the number of cues acquired (i.e. the number of tests participants chose to 'run' on the samples). The figure shows that participants tended to 'widen the jaws' or increase evidence accumulation from block 1 to block 2. This increase is perhaps not surprising given that the RAT strategy leads to more correct inferences than TTB in block 2. More surprising is the *decrease* in evidence accumulation observed in block 3. Here, participants learn to 'narrow the jaws' again even though the accuracy of the TTB and RAT strategies is identical in block 3. The differences in cue acquisition are small but the pattern in both experiments led to significant quadratic trends, F(1, 29) = 30.58, p < .001 and F(1, 28) = 35.41, p < .001 for Experiments 1 and 2 respectively, reflecting the upturn from block 1 to 2 and then downturn from 2 to 3.

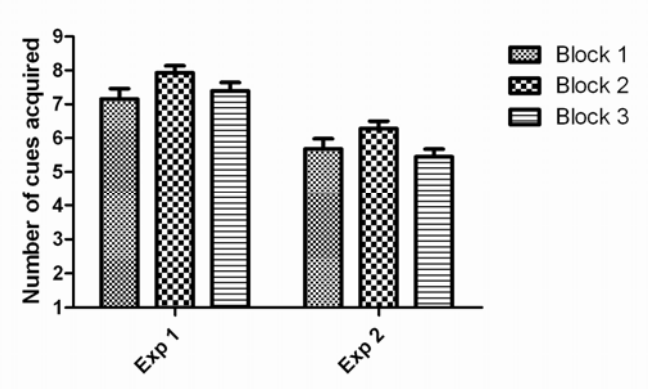

Figure 3: The number of cues acquired in each block of Experiments 1 and 2.

Figure 4 shows that this tendency to decrease evidence accumulation in block 3 relative to block 2 was present in the clear majority of individuals. The figure uses data from Experiment 2 and plots the average difference in the number of cues acquired in the last 50 trials of block 2 and the 50 trials of block 3. If participants decrease their search this value is positive; if they increase it is negative. The figure shows that 24/29 (83%) participants had a positive value. In Experiment 1 (not plotted) the proportion was also 83% (25/30 participants).

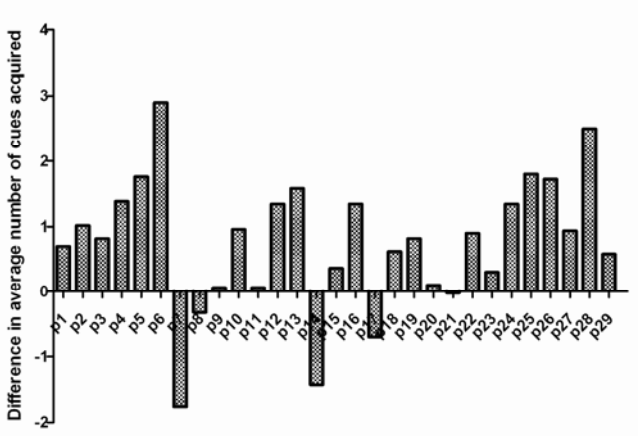

Figure 4: Individual data showing the difference in the average number of cues acquired in the last 50 trials of block 2 and the 50 trials of block 3 (Experiment 2).

A final measure of cue acquisition, anticipated in Figure 1, examined the number of cues acquired beyond the single discriminating cue predicted by the TTB strategy, as a proportion of how many cues remained. This measure is important because on some trials several cues need to be examined before a discriminating one is found; a fact that is not taken into account when only the raw number of cues acquired is considered. To illustrate: if the TTB 'stopping point' on a given trial was 3 cues and a participant acquired 4 of the 6 remaining cues, a value of .75 (4/6) would be recorded for the 'extra cues' measure. Figure 5 shows that the acquisition of extra cues follows the now characteristic pattern of an increase from block 1 to 2 and a decrease from blocks 2 to 3. The lower proportions over-all in Experiment 2 presumably reflect the additional opportunity cost of the time manipulation. The quadratic trends were highly significant in both experiments $F(1, 29) = 32.08$, $p < .001$ and $F(1, 28) = 49.51$, $p < .001$, for Experiments 1 and 2 respectively.

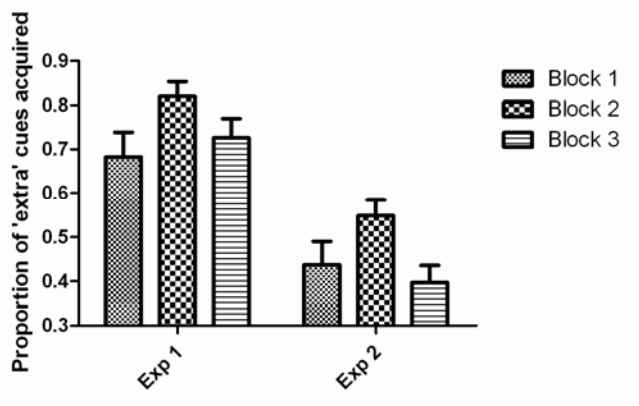

Figure 5: The number of extra cues (i.e., those beyond a single discriminating cue) acquired, expressed as a proportion of the remaining cues in each block of Experiments 1 and 2.

Discussion

In two experiments, participants increased their evidence accumulation when the environment changed to one in which a strategy requiring more information performed more accurately, and subsequently decreased accumulation when the environment changed back to one favoring a more frugal information search. Importantly, although the first change in the environment was signalled by a change in the *accuracy* of competing strategies, the second change could only be detected through the exploration of alternative levels of evidence, because both strategies were matched in terms of accuracy. These results contrast with previous failures to induce shifts in the strategies people adopt for making decisions in dynamic environments (Bröder & Schiffer, 2006; Rieskamp, 2006). The contrasting results are perhaps due to the different way in which adaptive changes are defined in this experiment and previous research. Earlier studies defined adaptive changes as a shift in the relative fit of a strategy (TTB or WADD) over blocks of trials; here a change is defined as shifts in the amount of evidence accumulated (trial-by-trial) following changes in the environment. It is possible that this latter definition increases the likelihood of observing adaptation.

The shifts in levels of evidence appear to have been driven by a desire to balance accuracy with the time cost of obtaining information. In this sense, the shifts in search behavior moving from block 2 to block 3 show that people are self-regulating their decision-making, and are not solely adapting on the basis of accuracy. The time delay in Experiment 2 had an impact on the overall levels of evidence accumulated but did not appear to accentuate the differences in evidence accumulated between blocks.

Many investigations of the adaptive nature of decision making pre-suppose that participants have access to a repertoire of cognitive strategies (Bröder & Newell, 2008; Payne et al., 1993; Rieskamp & Otto, 2006). Strategies are then selected according to the constraints of the environment, and presumably the preferences of the individual. In many recent studies the 'selection problem' has been reduced to one between TTB-like strategies and WADD- or RAT-like strategies. Our theoretical perspective differs from this in arguing that such behaviors are extremes in a sequential sampling evidence accumulation model (Lee & Cummins, 2004; Newell, 2005). We believe that the current experiments provide further support for this perspective. Note that strict adoption of a TTB strategy entails stopping search as soon as a single discriminating cue is found (Gigerenzer & Goldstein, 1996); such behavior would lead to a value of 0 Figure 5. Clearly very few participants adopt this strict form of TTB as the average proportion of extra-cues considered ranges between approximately 0.4 and 0.75 even in those environments in which TTB performs well (blocks 1 and 3). Similarly, a strict RAT strategy predicts the accumulation of *all* cues on every trial; the values displayed in Figure 3 shows that, on average, such

behavior was not observed. Thus, shifts in evidence accumulation can be interpreted not as transitions between discrete strategies, but shifts in a continuum of evidence.

The parsimony of such an explanation relies on an adequate model of how people learn to regulate their evidence threshold. In addition, preferring such an explanation requires situations in which an evidence accumulation model can provide a better account of behavior than a strategy selection model. Such work is yet to be done, but we can speculate about how successful the two approaches might be given the current data sets. The SSL model of Rieskamp and Otto (2006) uses a reinforcement mechanism to update the *expectancy* of a given strategy. It is able to use both *accuracy* and 'effort' signals to update expectancies. For example, in Study 3 of Rieskamp and Otto the model was able to capture participants' transition towards TTB in an environment where TTB and WADD made approximately the same number of correct predictions but information was costly, thus favoring a TTB strategy. Nevertheless, SSL predicts *inertia* effects when environments change (the environment was constant in Rieskamp and Otto Study 3) and thus it might have difficulty capturing the relatively fast transition between levels of evidence often seen in our data. The ability of the model might depend also on how adaptive change is defined (see earlier).

Another candidate model of threshold regulation is the Self Regulating Accumulator model developed by Vickers (1979). This is a sequential sampling model that uses 'boundaries', corresponding to levels of evidence, which control how much information is gathered before a decision is made. It also proposes mechanisms that adjusts these boundaries on a trial-to-trial basis, and so provides an account of learning and adaptation. Crucially for our data, a large part of this adaptation is *self*-regulation, based not on external feedback, but on controlling the internal level of confidence the model has in its decisions. This capability would explain the shift in search behavior moving from block 2 to block 3 in our experiment, and the way that Vickers (1979) proposed the boundaries are adjusted would also potentially predict the relatively sudden shift in search behavior we observed.

In conclusion, we believe our results present clear guidance and challenges for understanding how information search is regulated in human decision-making. While previous research has emphasized accuracy as a basis for adaptation, our results suggest this alone cannot be sufficient, and some form of internal self-regulation is also important. Possible theoretical ideas for understanding self-regulation include the notion of minimal effort and adaptation based on controlling internal levels of decision confidence. We plan to pursue these ideas to develop models of how people adapt their search and decision-making in changing environments and circumstances.

Acknowledgments

This research was supported by Australian Research Council Discovery Project Grants (DP 0770292; DP 0877510) to BRN and Air Force Office of Scientific Research grant AFOSR FA9550-07-1-0082 to MDL.

References

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review, 3*, 439-449.

Bröder, A. & Newell, B.R. (2008). Challenging some common beliefs about cognitive costs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making 3*, 205-214.

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology Learning, Memory, & Cognition*, *32*, 904–918.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and "rational" models. *Psychonomic Bulletin & Review*, *11*(2),343–352.

Newell, B. R. (2005). Re-visions of rationality. *Trends in Cognitive Sciences*, *9*(1), 11–15.

Newell, B.R., Lagnado, D.A., & Shanks, D.R. (2007). *Straight Choices: The Psychology of Decision Making.* Hove, UK: Psychology Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64-99). New York: Appleton-Century-Crofts.

Rieskamp, J. (2006). Perspectives of probabilistic inferences: reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology Learning, memory & cognition, 32*, 1335-1370.

Rieskamp, J., & Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Journal of Experimental Psychology: General, 135*, 207-236.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *1960 IRE WESCON Convention Record, Pt. 4*, 96-104.

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review, 12*, 387-402.

Young, M. E., & Wasserman, E. A. (2005). Theories of learning. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (pp.161–182). London: Sage.