UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Adjusting the Spanner: Testing an Evidence Accumulation Model of Decision Making

Permalink

https://escholarship.org/uc/item/2jc250xt

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN

1069-7977

Authors

Newell, Ben R.
Collins, Patrick
Lee, Micheal D.

Publication Date

2007

Peer reviewed

Adjusting the Spanner: Testing an Evidence Accumulation Model of Decision Making

Ben R. Newell (ben.newell@unsw.edu.au) and **Patrick Collins (patrickc@unsw.edu.au)**
School of Psychology, University of New South Wales
Sydney, NSW, 2052, Australia

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

Abstract

An experiment examined two aspects of performance in a multi-attribute inference task: i) the effect of stimulus presentation format (image or text) on the adoption of decision strategies; and ii) the ability of an evidence accumulation model, which unifies take-the-best (TTB) and rational (RAT) strategies, to explain participants' judgments. Presentation format had no significant effect on strategy adoption at a group level. Individual level analysis revealed large intra-participant consistency, including some participants who consistently changed the amount of evidence considered for a decision as a function of format, but wide inter-participant differences. A unified model captured these individual differences and was preferred to the TTB or RAT models on the basis of the minimum description length model selection criterion.

Keywords: Decision-making; Multi-cue inference; Heuristics; Take the best; Sequential sampling

The idea that decision makers have at their disposal a variety of 'tools' or strategies that can be selected for particular tasks has proved popular in the literature (Gigerenzer & Todd, 1999; Payne, Bettman, & Johnson, 1990; Rieskamp & Otto, 2006). The 'adaptive toolbox' metaphor proposed by Gigerenzer and colleagues is a good example of such an approach. Proponents suggest that decision makers have access to a "collection of specialized cognitive mechanisms that evolution has built into the mind for specific domains of inference and reasoning" (Gigerenzer & Todd, 1999, p. 30). One of the key mechanisms or heuristics in this toolbox is the 'take-the-best' algorithm (TTB), a heuristic for choosing between two alternatives. The defining feature of TTB is that it terminates information search once a single cue that discriminates between alternatives has been discovered. In this sense, TTB differs markedly from 'rational' decision models that advocate complete information search and optimal weighting of information.

Despite the impressive success of TTB in simulation studies, such as its ability to perform well against computationally intensive models (Gigerenzer & Goldstein, 1996), empirical studies seeking evidence that participants adopt TTB are more equivocal (Bröder, 2000; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). In many experiments the results suggest that some people make choices consistent with TTB some of the time but a significant proportion of participants adopt strategies that violate all or some of TTB's rules, especially the 'single discriminating cue' stopping rule.
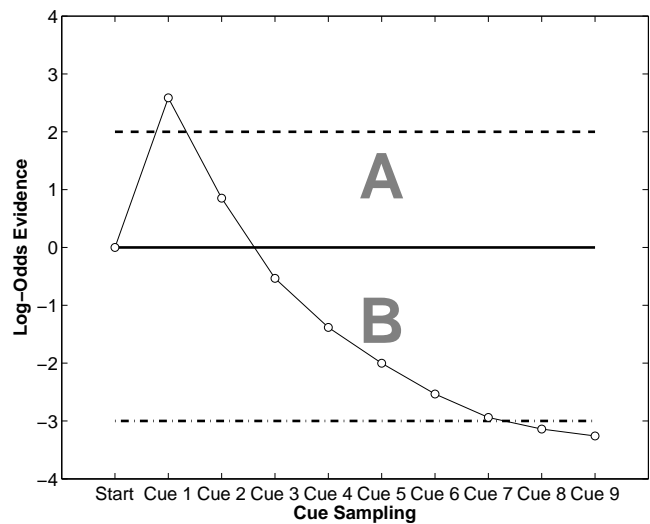

Figure 1: The unified sequential sampling model, showing the accumulated evidence as nine cues are sampled in validity order, and TTB-consistent (top) and RAT-consistent (bottom) decision thresholds.

In an attempt to account for this wide individual variability Newell (2005) suggested an alternative metaphor—an adjustable spanner (or wrench)—in which the width of the jaws represents the amount of evidence a person accumulates before making a decision. The important feature of an evidence accumulation model for two-alternative choice problems is that it can mimic the performance of TTB's stopping rule, or a strategy that incorporates more evidence, by adjusting the evidence required before a decision is made. Thus, one way of explaining individual variability is to suggest that all participants use an evidence-accumulation model but that some require greater amounts of evidence than others before making their decisions. Another appealing aspect of this model is that a 'single tool' circumvents the thorny issue of tool selection (cf. Rieskamp & Otto, 2006).

Lee and Cummins (2004) presented a formal instantiation of such an evidence accumulation model, which they proposed as a unification of TTB and rational models. As shown

in Figure 1, their approach was to view TTB and rational (RAT) as sequential sampling models. In sequential sampling models, information is accumulated as cues are observed, and a decision is made as soon as there is a threshold amount of evidence in favor of one alternative. Figure 1 gives an example where the first cue provides strong evidence (measured on a standard log-odds scale) in favor of decision A, but all of the subsequent lower validity cues favor decision B. Once all cues have been observed, there is more evidence for decision B than A. Accordingly, for low thresholds (the value two is shown as a concrete example) decision A will be made; for higher threshold values (the value three is shown as a concrete example) decision B will be made.

In general, low thresholds that guarantee sampling terminates as soon as evidence favoring one option is found will model TTB decisions, while high thresholds that guarantee exhaustive sampling of all cues will model RAT decisions. Thus, the unified model views these alternatives as special cases of a single evidence accumulation model corresponding to low (TTB) and high (RAT) evidence thresholds. In a multiple-cue judgment experiment, Lee and Cummins (2004) found that the unified model accounted for the highest proportion of participants' decisions (84.5%) and was favored by a minimum description length (MDL) model selection criterion sensitive to the additional complexity of the unified model.

An important next step in exploring the capability of a unified model to describe people's judgments is to specify how the evidence threshold is affected by factors such as the consequences and utility of decisions, and the supply and availability of information in the external environment (Lee & Cummins, 2004; Newell, 2005). The current experiment seeks evidence concerning the latter by examining a variable—the format by which stimulus information is presented—that has been shown to impact on the adoption of decision strategies.

Effects of Stimulus Format

Previous Findings

Bröder and Schiffer (2003, 2006b) presented participants with a multiple-cue judgment task in which the aim was to identify the perpetrator of a crime. In the learning phase information about the clothing worn by potential suspects was presented either as text descriptions (e.g., "green shirt") or as schematic images of people in different outfits. In a test phase the names of pairs of suspects were presented and participants had to decide which had the higher probability of being the perpetrator. Thus, at test, participants had to retrieve cue information learned previously from memory.

The key finding was that participants who had learned cue information from images tended to rely on RAT-type strategies at test whereas those who had learned cue information from text tended to rely on TTB. Bröder and Schiffer (2003, 2006b) interpreted the format effect in terms of the higher cognitive costs involved in retrieving text information from

memory relative to image information. Images present cue information as an integrated whole and so are perhaps more likely to be retrieved as such. Text lists are discrete and so conceivably features are retrieved sequentially, and so are perhaps suited to TTB with its single discriminating cue stopping rule.

Evidence concerning the format effect in more typical 'inferences from givens' tasks, in which cue information is presented visually rather than having to be retrieved from memory, is less clear. Juslin, Olsson, and Olsson (2003) found no differences in the adoption of TTB-consistent strategies when using text or image cue presentation (in both conditions there was very little evidence for TTB); although other aspects of their design made it less than ideal for testing TTB. On the other hand, Bergert and Nosofsky (2007) found considerable support for their 'generalized TTB' model in an experiment in which cue information was presented as integrated images. Such contrasting findings suggest that systematic research is required to understand how format affects the adoption of decision strategies.

Current Experiment

To investigate these issues, the current experiment involves a multiple-cue judgment task in which cue information was presented either in text or image format (between-subjects). Feedback was provided to enable learning, but the cue environment was constructed such that neither the RAT nor TTB model would be favored. In the subsequent test phase participants were given pairs of alternatives for which RAT and TTB made different predictions. Test items were presented separately in text and image formats (within-subjects).

If the switch between different strategies is driven by the format of the information, as per Bröder and Schiffer (2003, 2006b), we expect to see a dominance of RAT choices in the image test phase and a dominance of TTB choices in the text test phase. Alternatively, participants might adopt the strategy best suited to their training regime (whether image or text) and then 'stick' with it regardless of potential changes in the costs of applying it (cf. Bröder & Schiffer, 2006a). In addition to these questions about the effect of format, our other main goal is to ask whether the unified model with its evidence threshold parameter can better account for decisions than the deterministic RAT and TTB models.

Experiment

Participants

Forty-eight undergraduate students from the University of New South Wales participated in the experiment in return for course credit. There was an error storing the data for one participant, giving a final total of 47 participants.

Stimuli

The experiment used the cue environment developed by Lee and Cummins (2004), who give a full description of its con-

Stim. No.	Cue 1 (.97)	Cue 2 (.90)	Cue 3 (.82)	Cue 4 (.64)	Cue 5 (.56)	Cue 6 (.55)	Decision Variable
1	0	0	0	1	0	0	16
2	0	1	0	0	1	0	18
3	0	0	1	0	0	1	21
4	0	0	0	1	1	0	25
5	0	0	0	0	1	0	31
6	1	0	0	0	1	1	40
7	0	0	1	1	1	1	44
8	1	1	0	1	0	0	51
9	1	1	1	0	0	1	62
10	1	1	0	0	1	0	70
11	1	1	0	1	1	1	97
12	1	1	1	1	0	0	104
13	1	1	1	1	1	1	280
14	1	1	1	1	0	1	285
15	1	1	1	0	1	0	347
16	1	1	1	1	1	0	444

Table 1: The stimulus environment, showing cue patterns, cue validites, and decision variable values.

Figure 2: Examples of the schematic image format for displaying stimuli.

Test Pair	TTB Stimulus						RAT Stimulus					
1	1	0	0	0	0	1	0	1	1	0	0	0
2	1	0	0	0	1	0	0	1	1	0	0	0
3	1	0	0	0	1	1	0	1	1	1	0	0
4	1	0	0	1	1	0	0	1	1	1	0	0
5	1	0	0	1	1	1	0	1	1	1	1	0

Table 2: Test stimulus pairs, showing the assignment of cues to the TTB- and RAT-consistent stimulus.

struction. The environment comprises 16 objects described by six binary cues. Table 1 displays this environment showing the cue patterns for each stimulus and the validity of each cue.

The cue environment was instantiated as six pieces of clothing—baseball cap, t-shirt, handbag, skirt, stockings and shoes—each of which could be one of two colors. The stimuli were either text descriptions of these clothing items, or schematic images of a woman with these items, as shown in Figure 2. The assignment of cue validities to clothing items was random and differed for each participant. The cues were substitutive rather than present or absent to enable the cover story to be implemented (see below; cf. Bergert & Nosofsky, 2007).

Procedure

Training Phase   Participants were told that they were an undercover agent in a fictional country and had to learn about the clothing characteristics of members of a secret society. In the training phase all but one possible pairings of the 16 objects were presented. The exception, as with Lee and Cummins (2004), is the pairing of the second and seventh stimuli. This pairing was omitted because the TTB and RAT models make opposing predictions. For all remaining 119 pairings, the TTB and RAT models make the same prediction.

On each trial in the training phase the two paired stimuli were presented on screen, and participants selected the woman they thought more likely to be a secret society member. Twenty-four participants were trained using the text descriptions of the people, and the remaining 23 participants were trained using the schematic image representation. In both formats of presentation, the correct answer was determined by the stimulus with the higher decision variable, and feedback was given on each trial by indicating the correct choice.

Test Phase   Following training, participants completed a test phase, involving two blocks of 20 trials. Both blocks were comprised of four repetitions of five test questions, described below, presented in a random order for each participant. One of these blocks of 20 trials was presented using the text format, while the other was presented using the image format. The order of the formats was counterbalanced across participants. No feedback was given during the test phase.

As detailed by Lee and Cummins (2004), the five repeated test questions involved stimulus pairings for which the TTB and RAT model make opposing predictions. These pairs are displayed in Table 2. For each pair, the TTB model selects the stimulus on the left because it has a positive value for the most predictive cue. The RAT model makes the opposite prediction because the stimulus on the right always has more evidence favoring it once all the cues are assessed.

Results

Training Phase   Figure 3 displays the proportion of correct decisions for the 119 training trials, divided into six blocks, as a function of the text and image stimulus formats. The
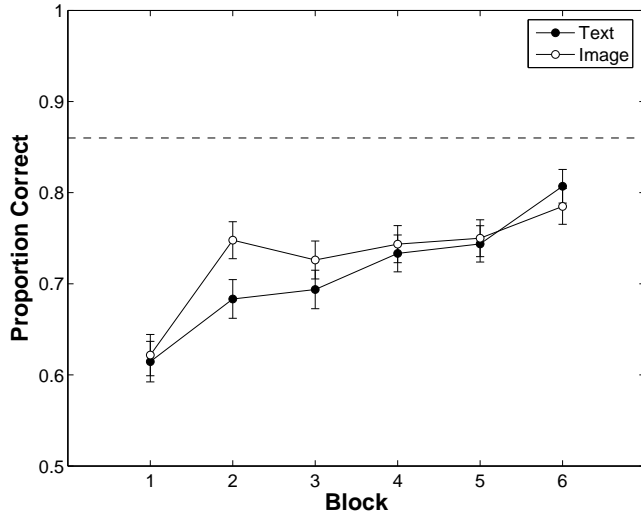
Figure 3: The mean proportion and standard error of correct predictions during training averaged across participants, for the text and image formats. The dashed horizontal line shows the theoretical maximum of 0.86.

	Train Text	Train Image
Test Text	0.54 (0.39)	0.50 (0.44)
Test Image	0.53 (0.44)	0.54 (0.39)

Table 3: The proportion (standard deviation) of TTB decisions across all participants as a function of training and test phase stimulus format.

Strategy	Number of Participants
TTB-Consistent	12
RAT-Consistent	8
Switch-Consistent	9
Inconsistent	18
TOTAL	47

Table 4: Number of participants classified as making decisions consistent with the different strategies.

clear increase in accuracy across blocks for both groups indicates that participants were able to learn from the feedback, and approaches the theoretical maximum of 0.86 shown by the horizontal dashed line. This theoretical maximum corresponds to the proportion of pairs for which the TTB and RAT models make correct predictions.

The proximity of the lines for the two groups indicates that stimulus format had little impact on learning. These two observations were confirmed by statistical analysis. There was a significant linear trend for Block, $F(1, 45) = 48.80$, $p < .001$, no effect of group, and no interaction between the two variables $Fs < 1$. The level of performance is comparable with other studies that have used the same cue environment (Bergert & Nosofsky, 2007; Lee & Cummins, 2004).

Test Phase For each test phase comparison a decision was either consistent with the prediction of the TTB model or consistent with the RAT model. There was no effect of the order of the text and image blocks at test. The effect on proportion of TTB-consistent decisions if the text block was first or second was $F(1, 45) = .04$, $p > .05$. The effect if the image block was first or second was $F(1, 45) = .35$, $p > .05$. Accordingly, the proportion of TTB- and RAT-consistent decisions was collapsed across both orderings.

Table 3 details the proportion of TTB-consistent decisions across all participants as a function of training and test stimulus format. At a group level there is no support for the prediction that TTB-consistent decisions dominate with the text format of presentation and RAT in the image format of presentation, with $F(1, 45) = .026$, $p > .05$. Neither is there any evidence the format used in training affects peformance on

the different formats at test, with $F(1, 45) = .009$, $p > .05$. Finally, there is no evidence of an interaction between the training and testing formats, with $F(1, 45) = .120$, $p > .05$. Indeed, remarkably, when collapsed across participants there are almost equal proportions of TTB- and RAT-consistent decisions regardless of training and test stimulus format.

The large standard deviations in Table 3 suggest, however, that collapsing across participants masks individual differences. As discussed by Lee and Cummins (2004), a more useful analysis considers decisions within rather than across participants. Accordingly, participants were classified as TTB-consistent, RAT-consistent, inconsistent with either model, or switch-consistent. The last of these were participants who consistently made decisions consistent with different models in the two test blocks. The criterion for consistency was that 80% of decisions were made following the prediction of the corresponding model[1]. Table 4 displays the number of participants classified according to each of these strategies.

Supporting the conclusions of earlier work (Lee & Cummins, 2004; Newell & Shanks, 2003; Newell et al., 2003), the data in Table 4 show clear inter-individual differences but strong intra-individual consistency in decision strategies. Over 60% of the sample was classified as consistently adopting one strategy throughout or switching between two consistent strategies in the two test phases. Of the nine switch-consistent participants, five made TTB-consistent decisions with the text format but RAT-consistent decisions with the image format, in line with the prediction from the work of Bröder and Schiffer (2003, 2006b). However, four participants showed the opposite pattern of behavior. Accordingly, no strong conclusions can be drawn regarding the motiva-

[1] The binomial probability of 16 out of 20 test phase responses in favor of one model is 0.998.

tion for switching. Strategy adoption appeared to be unaffected by training stimulus format with very similar numbers of TTB-consistent, RAT-consistent and inconsistent participants in both the image and text trained groups.

Unified Model Analysis

Our unified model analysis considers how the sequential sampling account best captures the TTB- and RAT-consistent decisions at the individual participant level. Relying on the results above, each participant's decisions were collapsed across training format and testing order.

There are four interesting modeling possibilities, each with clear theoretical interpretations, for these decision data. Under a pure *TTB Model* or *RAT Model*, both text and image decision-making follows solely the TTB or RAT model. Under a *Stable Individual Differences Model*, both text and image decision-making follow the unified model, with participants assigned to groups based on whether they have a TTB- or RAT-consistent majority of decisions. Both groups of participants then have their own parameterization under the unified model. Under a *Switching Individual Differences Model*, both text and image decision-making follow the unified model, with separate groups determined for both test decision-making phases. That is, TTB- and RAT-consistent majority groups are determined independently in each test phase. All four of these participant groups have their own unified model parameterization.

Clearly, the two individual differences models are more complicated than the parameter-free TTB and RAT models, and the individual differences model that allows for switching is more complicated than the stable individual differences model. It is important that these differences in model complexity be taken into account when fitting the models to data. To achieve this, we used the Minimum Description Length (MDL) model selection criterion, which is sensitive to both goodness-of-fit and model complexity, and was calculated using the 'entropification' method described by Lee (2004).

The results of the modeling analysis are detailed in Table 5, which shows the proportion of participants' decisions explained at the best-fitting parameterization of each model, and the MDL value. Lower values of the MDL criterion are better, and so the switching individual differences model, with an 87% accuracy, is clearly preferred. The other individual differences model is also clearly superior to either the TTB or RAT models.

This pattern of results reinforces two key findings. First, there are individual differences in the decision-making of participants. The MDL evaluation shows the patterns of inter-individual differences and intra-individual consistency are important regularities, not captured by the simpler pure TTB and RAT accounts. Secondly, there are patterns of switching behavior, with the same participants making TTB-consistent decisions in one presentation format, and RAT-consistent decisions in the other. The additional complexity of the switch-

Model	Accuracy	MDL
TTB	53%	191
RAT	46%	207
Stable	70%	155
Switch	87%	91

Table 5: Accuracy and Minimum Description Length (MDL) measures for the four models.

ing over the stable model for individual differences is needed to explain these differences. Thus, the superiority of the switching model identified by the MDL analysis shows that the participants who behaved this way represent an important regularity in decision-making on this task.

General Discussion

The experiment sought evidence concerning two aspects of multi-attribute decision-making: i) the influence on human judgments of the format in which stimulus information is presented; and ii) the capability of a unified decision model to describe human judgments. We consider each in turn.

Following the work of Bröder and Schiffer (2003, 2006b) we tested the hypothesis that judgments made about image stimuli would conform to more integrative 'rational' decision strategies and judgments made about text stimuli would conform to one-reason decision heuristics like TTB. We found no support for this hypothesis. At a group level, stimulus format exerted no systematic effect on the decision strategy adopted.

The key difference between our study and those of Bröder and Schiffer is that our participants had information presented visually, and so did not have to retrieve stimulus information from memory during the test phase. It appears that the format effect is dependent on the use of inferences from memory tasks. This is plausible given what is known about the nature of information stored in memory in pictorial and verbal format (e.g., Paivio, 1991). The benefit to more cognitively complex strategies (e.g., RAT) of having information integrated into a holistic visual representation is only conferred when the representation needs to be actively retrieved from memory. When the information is present at the time of the judgment it appears not to matter whether it is image or text based: either type can support either type of decision strategy. This absence of a format effect in memory from givens tasks is consistent with the findings of Juslin et al. (2003). In addition, finding that a substantial proportion of TTB-Consistent decisions were made with image-based stimuli is consistent with recent work by Bergert and Nosofsky (2007).

Perhaps the most illuminating aspect of the results is that, as with many previous studies, there is considerable evidence for inter-individual differences but intra-individual consistency. Newell (2005) has argued that such a pattern of results is problematic for a framework like the adaptive toolbox because of its assumption that the environment and not the indi-

vidual is the primary driver of strategy selection (Gigerenzer & Todd, 1999). To explain why people with (presumably) the same cognitive apparatus use different strategies in the same environment, the toolbox approach needs to posit multiple heuristics for one environment, which seems at odds with the thrust of the 'ecological rationality' argument.

Our modeling analysis suggests an alternative interpretation that is, perhaps, more appealing and parsimonious. Rather than positing multiple heuristics, the unified model suggests all participants use a sequential sampling process that includes TTB and RAT as special cases. We tested four models and found the best account was a unified model that allowed participants to switch between TTB- and RAT-consistent strategies in the different presentation formats.

When the results from the behavioral and modeling data are considered together an intriguing picture emerges. Although presentation format does not exert the systematic effect hypothesized from work on inferences from memory (Bröder & Schiffer, 2003), it is clear from the presence of the 'switch consistent' participants that for some individuals format does affect decision making. The existence of 'switch' participants is also interesting given recent findings regarding the routinization of decision strategies in multi-attribute tasks. Bröder and Schiffer (2006a) reported that many participants in their experiments adopted a particular strategy, and retained it regardless of environment changes that rendered the strategy maladaptive. At least some of our participants showed greater flexibility, and were able to adjust the evidence required for a decision as a function of a presentation format change in the environment.

Determining the motivation for switching, and potential characteristics underlying the individual differences in these tasks remains an avenue for future research. Future work should also focus on development of the unified model. Bergert and Nosofsky (2007) have argued that some of the assumptions of the model are too strong, in particular the assumed perfect knowledge of cue validities, making it interesting to examine tasks with active cue-discovery in the training phase. This would facilitate modeling the construction of cue-hierarchies and the cue-search process in general (cf. Rakow, Newell, Fayers, & Hersby, 2005).

Acknowledgments

References

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*(1), 107–129.

Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1332–1346.

Bröder, A., & Schiffer, S. (2003). "Take-the-best" versus simultaneous feature matching: Probabilistic inferences from memory and the effects of representation format. *Journal of Experimental Psychology: General*, *132*, 277–293.

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

Bröder, A., & Schiffer, S. (2006b). Stimulus format and working memory in fast and frugal strategy selection. *Journal of Behavioral Decision Making*, *19*, 361–380.

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

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Juslin, P., Olsson, H., & Olsson, A. C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156.

Lee, M. D. (2004). An efficient method for the minimum description length evaluation of cognitive models. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual conference of the cognitive science society* (pp. 807–812). Mahwah, NJ: Erlbaum.

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., & Shanks, D. R. (2003). Take-the-best or look at the rest? Factors influencing 'one-reason' decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 53–65.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, *91*, 82–96.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). *The adaptive decision maker*. New York: Cambridge University Press.

Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgment. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*, 1088–1104.

Rieskamp, J., & Otto, P. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236.