

In press. *Psychonomic Bulletin and Review*.

Sequential processing of cues in memory-based multi-attribute decisions

Arndt Bröder

Max-Planck-Institute for Research on Collective Goods, Bonn

Wolfgang Gaissmaier

Max-Planck-Institute for Human Development, Berlin

Corresponding author's address:

Dr. Arndt Bröder

Max-Planck-Institute for Research on Collective Goods

Kurt-Schumacher-Strasse 10

D-53113 Bonn, Germany

Phone: (+49) 228-916 14 53

Fax: (+49) 228-916 14 55

E-Mail: broeder@coll.mpg.de

KEYWORDS: Decision making; Simple Heuristics; Bounded Rationality; Memory search

Word count: 3,943 (excluding title, abstract, tables, figures, including references and notes)

ABSTRACT

When probabilistic inferences have to be made from cue values stored in long term memory, many participants appear to use fast and frugal heuristics such as "Take The Best" (TTB) that assume sequential search of cues. A simultaneous global matching process with cue weights chosen appropriately would mimic the decision outcomes, albeit assuming different cognitive processes. We present a reanalysis of response times from five published experiments ($n = 415$) and one new experiment ($n = 82$) that support the assumption of sequential search. In all instances in which decision outcomes indicated the use of TTB's decision rule, decision times increased monotonically with the number of cues that had to be searched in memory. Furthermore, response time patterns fitted the outcome-based strategy classifications which further validates both measures.

Sequential processing of cues in memory-based multi-attribute decisions

Although many decisions in real life depend on information we have stored in our long-term memory, decision rules have mainly been investigated in environments with information supplied by the experimenter, for example on the computer screen (e.g. Bröder, 2000b; 2003; Newell & Shanks, 2003; Newell, Weston & Shanks, 2003; Payne, Bettman & Johnson, 1993, but see Bröder & Schiffer, 2003b; Juslin, Olsson & Olsson, 2003). However, Gigerenzer and Todd (1999) speculated that these "inferences from givens" form an exception in everyday life which is more dependent on "inferences from memory". In addition, they argued that due to retrieval costs, inferences involving memory search would promote the use of noncompensatory fast and frugal heuristics, an example of which is "Take The Best" (TTB, Gigerenzer & Goldstein, 1996). In deciding between options, TTB goes with the most valid piece of information ("cue"), ignoring all others. People's decisions can sometimes be described with TTB, especially if information search is constrained by high costs or time pressure (e.g. Bröder, 2000b; 2003; Newell & Shanks, 2003; Newell et al., 2003; Payne et al., 1993; Rieskamp & Hoffrage, 1999).

In contrast to the usual screen-based research paradigm, Bröder and Schiffer (2003b, in press) implemented the idea of memory search in cue based decisions by introducing a cue-learning paradigm in which participants acquired knowledge about cues describing objects. The results broadly confirmed the claim that memory-based decisions are often noncompensatory and can be described by TTB for most of the participants.

However, when people use information from memory rather than from the screen, it is impossible to observe how they actually search for information. The inability to observe information search (called *process tracing*) poses a methodological challenge because one can only rely on outcome-based measures to decide which decision strategy someone is apparently applying. On the mere outcome level, compensatory procedures could produce decisions

indistinguishable from noncompensatory strategies if the dimension weights are chosen appropriately (Martignon & Hoffrage, 2002). In addition, further evidence is needed to distinguish models assuming sequential feature processing (such as fast and frugal heuristics) from models assuming a global matching process, such as Image Theory (Mitchell & Beach, 1990), Minerva-DM (Dougherty, Gettys, & Ogden, 1999), or PROBEX (Juslin & Persson, 2002). Global matching models assume that a probe is compared to all information in memory resulting in an activation depending on the similarity of probe and stored information. Although the models differ in their details of describing the feature-based similarity match, the process appears as a simultaneous assessment rather than sequential feature comparison.

Recently, Bergert and Nosofsky (in press) analyzed response times as convergent evidence for an outcome-based strategy classification in a decision making task from givens. We suggest that response time analyses could similarly be applied to investigate different strategies in memory based decisions. We hypothesize that response times increase with the number of information pieces that have to be retrieved to make a decision, which differs for different items and/or strategies.

We will report response time analyses of 5 published experiments and one new experiment.

Experiments 1 - 5

All experiments reported subsequently differed in minor procedural details which are fully reported in Bröder and Schiffer (2003b; in press). All studies employed a hypothetical criminal case involving 10 suspects of a murder: A famous singer was murdered near the pool, presumably by one of his former girlfriends. The participants were asked to help find the murderer. The basic idea of all the studies was to separate the acquisition of knowledge about the

suspects from making decisions about them, so that knowledge had to be retrieved from memory when making decisions.

Each experiment consisted of four phases: First, in an anticipation learning paradigm, participants acquired knowledge about the individual cue patterns of 10 suspects, which differed on four cues (e.g., dog breed). Each of the cues could have three different values (e.g., Spaniel, Dalmatian, or Dachshund). A portrait and a name of a suspect appeared on the screen, and participants had to reproduce the cue values with appropriate feedback. All 10 patterns were repeated until 90% of the responses were correct, indicating a sufficiently reliable knowledge base in memory.

To avoid participants' making inferences already during learning, a cue hierarchy was established only in a second phase by informing them about the evidence (cues) witnessed at the site of crime and about its relative importance. The relative importance of the four cues (predictive *cue validity*) was established by telling participants how many witnesses agreed on them. For example they were told that four witnesses agreed that the suspect had a Spaniel dog, whereas only two witnesses agreed that the suspect was wearing leather trousers.

The third phase consisted of complete paired comparisons of all suspects in which participants had to decide which suspect was more likely to be the murderer. Importantly, only the name of the suspects and their portrait were displayed. The cues that allowed deciding between the two suspects had to be retrieved from memory.

After this decision phase, a final memory test assessed the stability of cue memory as a manipulation check.

Description of the Strategies and Response Time Predictions.

The strategies we considered to potentially underlie the participants' decisions are TTB, Dawes's Rule (DR), Franklin's Rule (FR), and guessing. When comparing two suspects, the lexicographic TTB heuristic assumes that participants sequentially retrieve cues describing the suspects in the order of their validity. A person using TTB searches the most valid cue for both suspects first. If this cue discriminates, the person does not search further and makes a decision. Otherwise, searching for cues (in order of validity) continues until a discriminating cue is found. Therefore, the best (i.e., most valid) discriminating cue determines when TTB stops searching and decides, so that we predict a monotonic increase in response times depending on the number of cues that have to be retrieved until this best discriminating cue is found.

DR, going back to Robyn Dawes's (1979) work on unit-weight models, is a rule that takes all cues into account but does not consider their validity. A DR user tries to retrieve all cues and decides for the suspect that is "favored" by more cues. In case of a tie, the person has to guess. FR, similar to DR, takes all the information into account, but weighs it according to cue validity. Since less valid cues can overrule more highly valid cues, DR and FR are *compensatory* strategies. Both DR and FR, at least in a strict sense, require searching all cues in an unspecified order. Response times should therefore not depend on which is the best discriminating cue. Since DR and FR do not specify a search order, it is more difficult to distinguish between sequential search and global matching for users of these strategies. One prediction that, in our view, follows from the sequential search assumption but not from the global matching assumption is that people classified as using FR should be slower on average than DR users, since FR, additionally to DR, requires weighing cues according to their validity and is thus cognitively more complex.

The last strategy, *guessing*, consists of retrieving no cues at all and just randomly decides for one of the suspects. Therefore, the response times of guessers should not vary with the position of best discriminating cue, and they should be the quickest overall.

Results and Discussion

To decide which of the strategies someone was apparently using, the choice vector produced by each participant was classified by a maximum-likelihood procedure, details of which are provided in Bröder and Schiffer (2003a). The method assumes a uniform response error probability and stable strategy use across trials and determines the strategy for which the data are most likely.

Table 1 contains an overview of the experiments reported previously. In sum, the results show that the need to retrieve cue information from memory induced fast and frugal decision making, especially when cues were represented verbally and when working memory load was high.

To analyze response times, we combined participants from all 5 experiments and split them into four groups with identical strategy classifications. There were 198 TTB users, 90 FR users, 83 DR users and 44 participants who appeared to guess. Nine unclassified patterns (with identical likelihoods for more than one strategy) were excluded. For each participant, we computed the outlier-robust median response time for each of the four item types, depending on the position of the best discriminating cue. These individual response time medians were entered in the subsequent ANOVA.¹

The mean response times for each strategy group are depicted in Figure 1. There was a main effect of the position of the best discriminating cue, Greenhouse-Geisser-corrected $F(2.53, 1041.48) = 20.63, p < .001$, showing increasing decision times in general. There was also a main effect of strategy, $F(3, 411) = 6.92, p < .001$ and, much more important, a significant interaction,

$F(2.53, 1041.84) = 3.41, p = .001$. For all strategies, the decision times increased with the position of the best discriminating cue, but this increase was most pronounced for TTB users. To substantiate this claim, regression slopes were computed for each individual and compared across strategy groups, showing an overall difference, $F(3,411) = 9.52, p < .001$. According to Scheffé posthoc tests, TTB slopes ($B = 1.31$) differed significantly from all others, whereas DR, FR, and GUESS slopes did not significantly differ (B 's = 0.49, 0.24, and 0.26, respectively). Neither did FR and GUESS slopes significantly differ from Zero, both $t(>42) < 1.30$, both $p > .20$, whereas TTB slopes did, $t(197) = 8.49, p < .001$, as did DR slopes, $t(82) = 3.27, p < .01$.

The response times thus followed the predicted pattern and supported the assumption of sequential search: The increase was much less pronounced for FR, DR, and guessing than for TTB which would be expected if people generally search more cues than the best discriminating one, or no cues at all (when guessing). Still, there is a slight descriptive increase in response times also for FR and DR users, which we will discuss in the General Discussion. FR users needed generally more time than DR users which we expected, given that DR users only have to count evidence, while FR users also have to weigh it. Participants with predominantly non-systematic guessing behavior generally needed less time than all others.

An alternative interpretation is that the results are not due to sequential search but to option similarity and hence item difficulty. The more nondiscriminating cues TTB has to retrieve, the more similar on average the options must be. Hence, both variables are confounded. If the sequentiality assumption is correct for TTB users, their decision times should be related more strongly to the position of the best differentiating cue than to the number of identical cue values indicating similarity or difficulty. Since position and similarity are correlated, we computed individual multiple regressions of response times on both predictors. As expected, people classified as using TTB showed a steeper increase with position (mean $B = 1.08$) than

with similarity ($B = 0.37$), $t(197) = 2.76$, $p < .01$, whereas there was the opposite tendency for DR users (.14 vs. .57), $t(82) = -1.69$, $p = .09$. There was no difference in slopes for FR users (0.26 vs. -.02, $t(89) = 0.89$, $p = .37$). The same pattern of results emerges when raw rather than partial correlations are analyzed by Wilcoxon tests or t-Tests on the Fisher-Z-transformed values. Hence, we find no support for the alternative explanation that the TTB response time results were caused by item difficulty rather than sequential search, whereas the opposite is true for DR users.

However, a second criticism could argue that the results reported here are an artifact of a procedural peculiarity used in Experiments 1 to 5: In all cases, the cue validity hierarchy was equivalent to the order in which the cue values were learned. Hence, the sequential search for cues in order of their validities could potentially reflect a search in order of learning. To rule out this interpretation, we conducted an experiment in which we disentangled both factors.

Experiment 6

Method

Design. Experiment 6 also used the paradigm of the invented murder case. A two group design was used to disentangle the two possible search orders, search by learning and search by validity. In the *match* condition, the cue validities matched the order in which the cues were learned. In the *mismatch* condition, cue order and validity order were different. For instance, the learning order was "dog breed, jacket, trousers, shirt color", whereas the scrambled validity order of cues was "trousers, dog breed, shirt color, jacket". Hence, if the cues are numbered according to the learning sequence, the validity order in the mismatch condition was 3-1-4-2. The labels of the cues were counterbalanced by reversing validity and learning order for half of the participants (i.e., "dog" was now the most valid cue and "trousers" was the topmost cue in the learning order). Both of those cue orders were also used in two counterbalanced match

conditions, in which learning and validity order coincided. Since there was no difference between the counterbalancing conditions, we merged them and subsequently only refer to cues in order of validity or in order of learning, irrespective of the actual content of the cue.

Participants. Ninety-four participants attended the study, almost 90% of them were students. Twelve participants were excluded because they did not reach the learning criterion in Phase 1 within one hour. The remaining 82 participants (50 female, age 25.6 years, $SD = 3.25$, 20 - 36) were randomly assigned to the match or the mismatch condition, resulting in 41 participants each. Participants received 15 € for their participation with an additional chance to win 10 €, granted to the 5 participants with the best performance in the final memory task.

Procedure. Participants first had to learn cue profiles of 10 potential suspects before a cue validity order was established. After that, participants had to compare pairs of suspects and decide which of the two was more likely to be the murderer. After the test phase, participants were asked again to indicate all cue values for all suspects in a final memory test.

Results and Discussion

Learning Phase. There was no difference between the conditions with regard to learning. People needed on average 85 trials to finish the learning phase ($SD = 20.43$; 65-142), and remembered 86% of cue values in the final test showing that cue knowledge was reliable.

Strategy Classification. People were classified as users of one of the different strategies following the outcome-based strategy classification after Bröder and Schiffer (2003a). In the match condition, cue validity order and learning order were identical. In the mismatch condition, however, participants may *either* use the validity order as a search sequence (according to TTB) *or* they search cues in the order of learning, which we will refer to as a "Take The First" (TTF) heuristic. Consequently, we included TTF in the set of possible strategies for the mismatch condition.

In the match condition, there were 21 TTB users, 9 FR users, 8 DR users, and 3 people who were guessing. Thus, like in the comparable Bröder & Schiffer (2003b) experiments, there was a majority of people whose decisions can best be described with TTB. In the mismatch condition, there were only 11 TTB users, 8 FR users, 12 DR users, and 5 people that appeared to guess. In addition, 5 of the 41 participants were classified as using TTF. The strategy distributions across conditions differed significantly, $\chi^2(4, N = 82) = 9.48, p = .05$. However, if TTF and TTB participants are combined into one class, the difference between the conditions disappears, $\chi^2(3, N = 82) = 2.04, p = .56$. Hence, the data support the interpretation that the proportion of TTB users found when validity and learning order are confounded may be a composite of "real" TTB users and others using TTF.

Decision times. There was no difference between the match and the mismatch condition regarding group decision times, $F(1, 80) = 0.06$, and no interaction between cue and condition, $F(2.17, 173.7) = 0.14$. Therefore, decision times for people classified as using the same strategy were pooled across the conditions. The mean decision times for each outcome-based strategy are presented in Figure 2. The pattern is very similar to the one we obtained when combining Experiments 1 to 5, although somewhat noisier since the sample size is only one fifth of the combined sample size of Experiments 1 to 5.

The main effect of cue is significant, $F(2.08, 160.19) = 6.55, p = .002$. Decision times increased with the position of the best discriminating cue. There was also a trend for a main effect of strategy, $F(4, 77) = 2.07, p = .09$, and an interaction between cue and strategy, $F(8.32, 160.19) = 2.40, p = .02$. Again, TTB users show the largest increase in response times, followed by FR, DR and Guessing. Response times are again generally higher for FR than for DR. A striking pattern can be seen for the five TTF users: They show the shortest response if Cue 3 discriminates, followed by Cue 1, Cue 4 and Cue 2. This pattern of response times exactly

matches the learning order of the cues. Note that this result is far from trivial: The strategy classifications were *exclusively* based on choices. Hence, this congruence of classification and response times constitutes true converging evidence for this strategy and the notion of sequential cue search.

General Discussion

The goal of this paper was to find converging evidence for processes we assumed to underlie memory-based multi-attribute decisions. Direct process tracing is not possible for memory-based decisions. Therefore, we analyzed response times to validate the idea of sequential cue search in multi-attribute decisions from memory and as an independent source of support for the outcome based strategy classification method.

Both the reanalysis of the 5 published experiments and the results of the new experiment support the idea of sequential cue search. For users of TTB this support is clearest, as their response times increased with the number of cues that need to be looked up until the best discriminating cue was found. This result seems better explained by sequential search than by global matching because regression analyses revealed only a weak relation between RT and item difficulty based on feature similarity. For users of DR and FR, the increase based on the position of the best discriminating cue was much weaker and, contrary to TTB users, depended more on item difficulty based on similarity (at least for DR). In principle, this could be explained by both sequential search and global matching. However, we think that the additional finding that DR users are generally quicker than FR users is more supportive of sequential search assuming that DR users only add up information, while FR users additionally weigh it. At least, it is not clear to us how a global matching model could explain this difference. Finally, for TTF users, response time patterns fitted the presumed search order and stopping rule, which also favors sequential search.

A slight increase in response times for users of compensatory strategies, as observed, is likely to occur nevertheless because it is possible that someone classified as using FR sometimes applies TTB. Moreover, retrieving information in order of validity also makes sense for compensatory strategies. For example, a FR user who knows that the two most valid cues point towards one suspect does not need to look up further information because the less valid cues could never overrule this judgment.

The strength of these sequential search effects surprised us. In none of the experiments were people instructed to decide quickly, which is usually a prerequisite to obtain interpretable response time data. Large inter- and intraindividual differences normally inflate noise and demand for large effects.

After demonstrating the usefulness of response time data in addition to outcome-based strategy classifications, the next step in evaluating the sequential search hypothesis should strive at reaching more precision in response times and classifications by moving from the group level to the individual level, which is the final appropriate benchmark for testing cognitive models.

References

- Bergert, F. B., & Nosofsky, R. M. (in press). *A response-time approach to comparing generalized rational and take-the-best model of decision making*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Bröder, A. (2000b). 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*, 1332–1346.
- Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *29*, 611–625.
- Bröder, A. & Schiffer, S. (2003a). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, *16*, 193–213.
- Bröder, A., & Schiffer, S. (2003b). Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, *132*, 277–293.
- Bröder, A. & Schiffer, S. (in press). *Stimulus format and working memory in fast and frugal strategy selection*. *Journal of Behavioral Decision Making*.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180 - 209.
- Gigerenzer, G. & Goldstein, D. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G. & Todd, P. M. (1999). Fast and frugal heuristics: the adaptive toolbox. In G.

- Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart*. (pp. 3–34). 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*, 133–156.
- Juslin, P., & Persson, M. (2002). PROBABILITIES from EXemplars (PROBEX): A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563–607.
- Martignon, L. & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*, 29–71.
- Mitchell, T. R., & Beach, L. R. (1990). ". . .Do I love thee? Let me count . . .": Toward an understanding of intuitive and automatic decision making. *Organizational Behavior and Human Decision Processes*, *47*, 1–20.
- 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, and 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.
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Rieskamp, J. & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.

Authors note

Arndt Bröder, Max-Planck-Institute for Research on Collective Goods; Wolfgang Gaissmaier; Max-Planck-Institute for Human Development.

The research was supported by the *Deutsche Forschungsgemeinschaft* (DFG, grant BR-2130/1-2) and by the Max-Planck-Institute for Human Development.

Address correspondence to broeder@coll.mpg.de.

We thank Thorsten Pachur and Lael J. Schooler for very fruitful comments on an earlier draft of this paper.

Footnotes

¹ Individual z-scores of response times or log-transformed response yielded the same patterns of significant results in all analyses.

Tables

Table 1

Overview of studies

Source		%TTB users x condition		N	Cue Descriptions
Bröder & Schiffer (2003b)	Exp. 1	load ^a	no load ^a	50	Blood type, cigarette brand,
		72.0 %	56.0 %		perfume, vehicle
	Exp. 2	memory	screen	50	Jacket, shoes, bag, vehicle
		44.0 %	20.0 %		
	Exp. 3	verbal	pictorial	50	Jacket, shoes, bag, vehicle
		64.0 %	64.0 %		
	Exp. 4	verbal	pictorial	114	dog breed, jacket, trousers,
		47.4 %	26.4 %		shirt color
Bröder & Schiffer (in press)	Exp. 5	verbal	pictorial	151	dog breed, jacket, trousers,
		69.7 %	36.0 %		shirt color
		load ^a	no load ^a		
		53.0 %	34.2 %		
This article	Exp. 6	match ^b	mismatch ^b	82	dog breed, jacket, trousers,
		51.2 %	26.8 %		

^a working memory load

^b between validity and ease of retrieval

Figure Captions

Figure 1. Mean decision times in seconds (and standard errors) as a function of best discriminating cue and outcome-based strategy classification in the combined Experiments 1 to 5.

Figure 2. Mean decision times in seconds (and standard errors) as a function of best discriminating cue and outcome-based strategy classification in Experiment 6 including five participants who apparently searched in the order of cue retrievability ("Take The First") in the mismatch condition.

Figures

Figure 1.

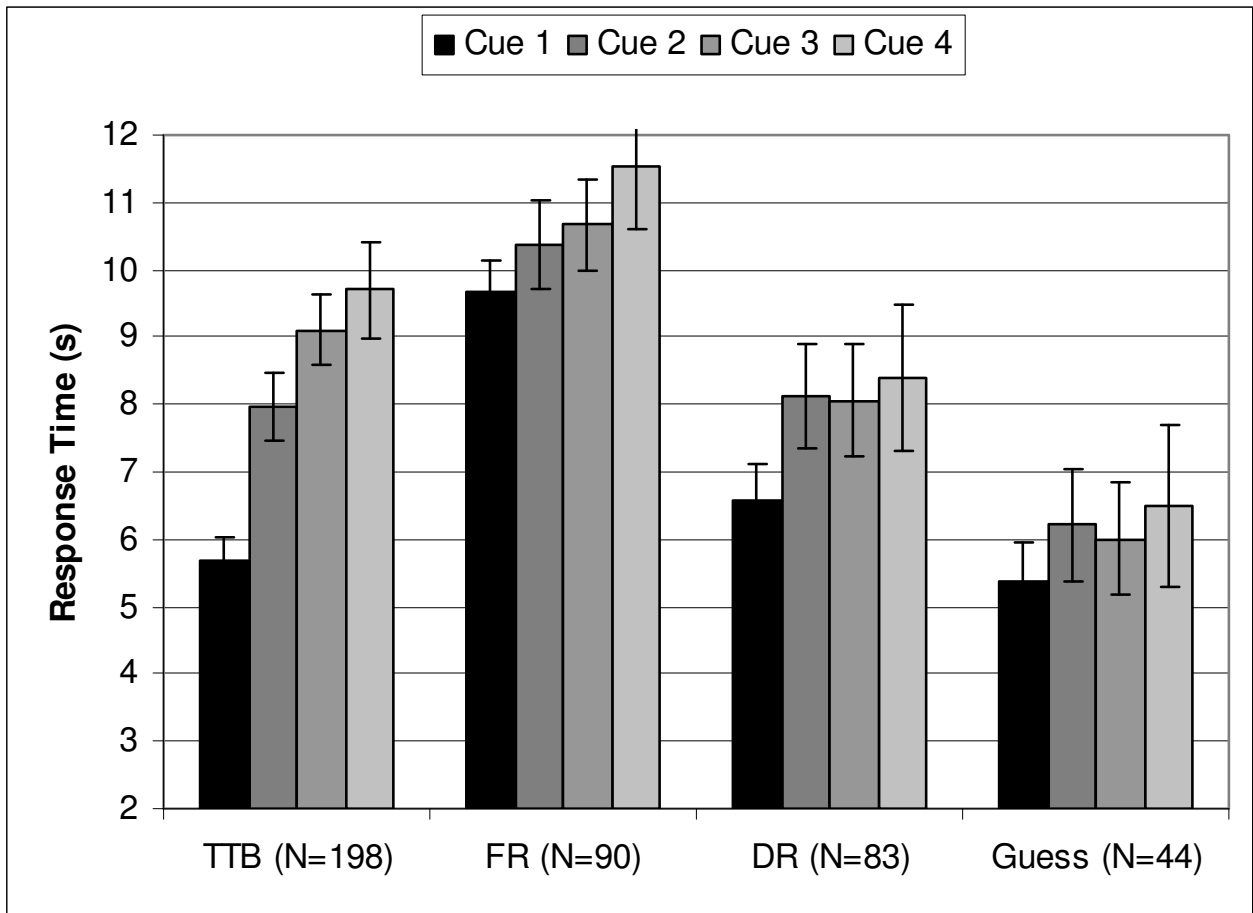


Figure 2.

