On the role of recognition in consumer choice: A model comparison

Benjamin E. Hilbig*

Abstract

One prominent model in the realm of memory-based judgments and decisions is the recognition heuristic. Under certain preconditions, it presumes that choices are based on recognition in a one-cue non-compensatory manner and that other information is ignored. This claim has been studied widely—and received, at best, mixed support—in probabilistic inferences. By contrast, only a small number of recent investigations have taken the RH to the realm of preferential decisions (i.e. consumer choice). So far, the conclusion has been that the RH cannot satisfactorily account for aggregate data patterns, but no fully specified alternative model has been demonstrated to provide a better account. Herein, the data from a recent consumer-choice study (Thoma & Williams, 2013) are re-analyzed with the outcome-based maximum-likelihood strategy classification method, thus testing several competing models on individual data. Results revealed that an alternative compensatory model (an equal weights strategy) accounted best for a larger number of datasets than the RH. Thereby, the findings further specify prior results and answer the call for comparative model testing on individual data that has been voiced repeatedly.

Keywords: model comparison, recognition heuristic, equal weights strategy, consumer choice, multinomial processing tree models.

1 Introduction

Many everyday judgments decisions must be made from memory. Correspondingly, a growing number of models in judgment and decision making explicitly consider the role of memory (Weber & Johnson, 2009). One such theory that has attracted substantial attention in recent research is the recognition heuristic (RH; Goldstein & Gigerenzer, 2002). The RH is a simple—and yet often surprisingly accurate—strategy proposed for comparative judgments from memory. It presumes that recognized options are chosen over unrecognized ones in a one-cue non-compensatory fashion, implying that no further information can overrule the binary all-or-none recognition cue (for the RH's preconditions, see Gigerenzer & Goldstein, 2011; Pachur, Bröder, & Marewski, 2008). The RH has been studied widely in the realm of probabilistic inferences (for reviews, see Gigerenzer & Goldstein, 2011; Pohl, 2011) and although some degree of controversy remains concerning the claim of one-cue non-compensatory decision-making (Brighton & Gigerenzer, 2011; Hilbig & Richter, 2011), there is consensus that it does account for the behavior of some individuals under certain circumstances (Hilbig, 2010; Pachur et al., 2008)—especially whenever the task induces the motivation to reduce cognitive effort (Hilbig, Erdfelder, & Pohl, 2012; Pohl, Erdfelder, Hilbig, Liebke, & Stahlberg, 2013).

Extending the RH theory beyond its original domain of probabilistic inferences, recent studies have investigated the role of recognition in preferential choice, that is, consumer decision-making (Oeusoonthornwattana & Shanks, 2010; Thoma & Williams, 2013). In both studies, participants were asked to choose between pairs of products. In critical pairs, a recognized brand was paired with an unknown one such that the RH would presume choice of the former. In addition, the authors manipulated whether additional information (i) was in line with the recognition cue (suggesting the same choice), (ii) was neutral, or (iii) contradicted the recognition cue (suggesting choice of the option with the unknown brand name). In line with previous research in probabilistic inferences (e.g. Bröder & Eichler, 2006; Newell & Fernandez, 2006; Newell & Shanks, 2004; Richter & Späth, 2006), the authors found that the probability of choosing recognized options—although substantial and well above chance level—was not independent of the additional information, thus contradicting the RH's claim of one-cue non-compensatory decision-making.

However, the result that aggregate choice probabilities are not perfectly aligned with the RH does not imply that another model must be preferred, let alone specify which one. Likewise, the individual analyses reported by Thoma and Williams (2013) hint only that the RH may have been used by some individuals, but probably not by others—while leaving open how these others may have been making choices. Correspondingly, Thoma and

*Department of Psychology, School of Social Sciences, University of Mannheim, Schloss Ehrenhof Ost, 68131 Mannheim, Germany. Email: hilbig@psychologie.uni-mannheim.de.

Table 1: Cue patterns for the three item types in Thoma and Williams' (2013) experiment and choice predictions of models.

Item type:	1 (positive		2 (neutral)		3 (negative)	
	A	B	A	B	A	B
Recognition	1	0	1	0	1	0
Additional information	1	0	–	–	0	1
Predictions:						
RH/WADD1		A		A		A
EQW		A		A		Guess
WADD2		A		A		B
GUESS		Guess		Guess		Guess

Williams (2013) conclude that "recognition is not used as the sole cue" (p. 42), but conclude that their results do not support any one specific alternative (compensatory) model ("the compensatory effect observed in this study is arguably not fully consistent with a simple cue integration model either", p. 42). Based on similar findings, some researchers have suggested that the RH be retained so long as no fully specified alternative model is shown to account for the data more successfully (Brighton & Gigerenzer, 2011; Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010; Pachur, 2011). However, even if a falsified model is retained simply because no better alternative is available, model refutation should undoubtedly trigger attempts to seek out and comparatively test potentially superior alternative models.

2   Methods

Despite the fact that comparative model tests are the exception in research on the RH (for a quintessential example, see Glöckner & Bröder, 2011) and have not yet been reported for preferential choice tasks, Thoma and Williams' data actually allow for such a test of competing models, since the degree to which additional information confirms or contradicts the RH was varied. In such a setup, the choice-based maximum-likelihood strategy classification method suggested by Bröder and colleagues (Bröder, 2010; Bröder & Schiffer, 2003) can be used. In essence, if models predict distinct choice vectors across two or more sets of trials (i.e. conditions), the models can be compared with respect to how well they fit the data. In Thoma and Williams' (2013) study, there are three such sets of trials (conditions) or *items types* as summarized in Table 1. Note that the strategy predictions and thus model compar-

ison hinge on the vital assumption that the strategy will be constant across all trials for each individual decision maker. Given that the nature of the information remains constant across trials (only its content varies), this is arguably a plausible assumption and one that is commonplace (Bröder & Schiffer, 2006a, 2006b; Glöckner, 2009; Glöckner & Bröder, 2011; Platzer & Bröder, 2012).[1]

As outlined above, in critical trials with exactly one option recognized, the additional information—in Thoma and Williams' (2013) study these were quality ratings—was either aligned with the recognition cue (item type 1), neutral (item type 2), or contradicted the recognition cue (item type 3). Participants made 10 choices in each of these item types and thus 30 in total. Across the three item types, several models make distinct choice predictions: The RH predicts choice of the recognized option in each of the item types—since any additional information should be ignored. An equal weights model (EQW; choose the option with the higher sum of positive cue values) predicts choice of the recognized option in item types 1 and 2, but it has to guess in item type 3 since the two cues contradict each other. A weighted additive model (WADD; choose the option with the higher sum of weighted cue values) will make different predictions depending on whether the recognition cue or the additional information receives more weight. If the recognition cue is given more weight (WADD1), the model is equivalent to the RH, thus making the same choice predictions. In turn, if the additional information is given more weight (WADD2), the model predicts choice of the recognized option in item types 1 and 2 and choice of the unrecognized option in item type 3.

To determine the "best" model for each participant, the vial assumption is made that the probability of choosing in line with a model's predictions is constant across item types (except when guessing is predicted in which case the choice probability is set to .50). Thus, models are allowed only unsystematic strategy execution errors, whereas systematic errors lead to misfit (Bröder & Schiffer, 2003; Moshagen & Hilbig, 2011). To determine model fit, individual choice frequencies of recognized options conditional upon item types were used. As Thoma and Williams (2013) thankfully published the raw data of their study, these could be calculated in a straightforward manner. Based on the choice frequencies, and resorting to the multinomial processing tree framework (Batchelder & Riefer, 1999; Erdfelder et al., 2009) the Bayesian Information Criterion (BIC, Wasserman, 2000) for each of the models per individual data set was determined using the freeware multiTree (Moshagen, 2010). Note that reliance on BIC was indicated since models differ in complexity (RH/WADD1, EQW, and WADD2 have on free parame-

---

[1]For an alternative approach and extension to strategy mixtures, see Davis-Stober and Brown (2011).

Table 2: Number of individual datasets (proportion of the total sample in parenthesis) for which competing models provided the best account (smallest BIC).

	Strict error criterion (.20)	Lenient error criterion (.40)
RH/WADD1	10 (16%)	19 (31%)
EQW	24 (39%)	24 (39%)
GUESS	16 (26%)	12 (19%)
WADD2	1 (2%)	2 (3%)
Unclassified*	11 (18%)	5 (8%)

* Datasets remain unclassified if all models are excluded, either due to absolute misfit (p < .05) or due to an observed error above the error criterion specified. The rationale is that these datasets were most likely generated by a decision strategy outside the set of those considered (Moshagen & Hilbig, 2011).

ter each whereas GUESS has none). Model equations can be found in Appendix A. The model which produced the smallest BIC was retained as the best description of an individual participants' data (for a similar approach in multinomial modeling see Hilbig, Erdfelder, & Pohl, 2011).[2] To rule out falsely classifying datasets which were most likely generated by some other, unknown process outside the set of models considered, only models that fit the data ($p > .05$) were retained for classification (Moshagen & Hilbig, 2011).

3   Results

The classification results are summarized in Table 2 for each of two different maximum levels of execution error (Bröder & Schiffer, 2003; Glöckner, 2009). The former implements a rather strict requirement, namely that models should predict choices very well and substantially better than chance level. The latter, in turn, is more lenient and requires only that models predict choices slightly better than chance. As can be seen, EQW provided the best account for the modal number of datasets. This result held regardless of the level of execution error that was allowed. Only a few datasets were best accounted for by RH/WADD1, especially with a stricter maximum level

of execution errors. Guessing also accounted for some datasets, whereas WADD2 accounted for essentially none.

As these findings reveal, EQW accounted for most participants' choices best. Note, however, that when EQW predicted guessing (i.e. in the negative condition) there were systematic item-level difference for participants classified as EQW-users. More specifically, whereas the aggregate probability of choosing the recognized brand was .52 and thus very close to chance, predicting a choice proportion of .50 (across participants) for *each* single brand led to misfit ($\chi^2(26) = 52.5$, $p = .002$). That is, the strict interpretation of EQW that guessing should occur for each item in the negative condition was rejected. Closer inspection revealed that this was primarily driven by two familiar brands which were extremely often chosen despite the additional negative cue, namely Apple (.82) and Toshiba (1.00). Once these two were excluded from the analysis, the strict EQW-hypothesis (choice proportion of .50 for each single brand across participants) fit the data well ($\chi^2(24) = 31.8$, $p = .13$).

To further assess the appropriateness of the classified models, participants' response times across the three item types were analyzed (for a similar approach, see Bröder & Gaissmaier, 2007). The RH predicts that response times should be equivalent across the three item types: As shown in Table 1, the recognition cue always discriminates and since choices should thus be based on this cue alone, response times should be constant across item types (Glöckner & Bröder, 2011). EQW predicts that all pieces of information are acquired and added in each of the item types, implying the same processing time[3]. However, in item type 3 an additional step is required, namely guessing (because the sum of cue values does not discriminate between options). Thus, according to EQW response times should be equivalent in item types 1 and 2, but longer in item type 3. Finally, GUESS predicts constant response latencies across all item types. WADD2 was not considered in the analysis since too few datasets conformed to this model in the choice-based classification.

For the analyses, the mean log-transformed response time was computed for each participant and item type. These means along with the classified model per individual data set (separately for the lenient and strict classification criterion) can be found in the online supplementary material. To test the above predictions, the mean response times were compared for each combination of item types (1 vs. 2, 2 vs. 3, and 1 vs. 3) using paired t-test, across all individuals classified into the same model category. Given that several null-hypotheses were tested (i.e., equivalent

---

[2]Prior to analyzing the actual data, a recovery simulation was run to ensure that the method would uncover the data-generating model reliably. To this end, 1000 datasets were generated under each of the models' choice predictions, assuming a strategy execution error of .10. Results revealed that 90% of datasets were classified correctly (the data-generating model produced the smallest BIC) and misclassifications were unsystematic, thus ruling out bias in favour of any one particular model.

[3]Note that in item type 2 the additional cue is not absent (this would imply less processing time), but simply does not discriminate between options.

—

Table 3: Mean differences (JZS Bayes factors in parenthesis) for each of the response time differences between item types, conditional upon strategy classification (lenient classification criterion). See Table 1 for item types.

	Items types tested against each other		
	1 vs. 2	2 vs. 3	1 vs. 3
RH/WADD1	.004 (5.6)	.041* (.53)	.045* (.49)
EQW	.020 (2.6)	.044* (.72)	.064** (.03)
GUESS	.019 (2.5)	.030 (3.0)	.048 (1.1)

Note: * $p < .05$, ** $p < .01$ (paired t-tests). Positive differences indicate that the former item type was responded to faster. JZS Bayes factors are odds in favor of the null-hypothesis (no difference); thus, values below 1 indicate evidence for the alternative hypothesis.

response times), the obtained t-values were transformed into JZS Bayes factors using the approach of Roulder, Speckman, Sun, Morey, and Iverson (2009). Thereby, the odds in favor of the null hypothesis (assuming uniform priors) were approximated. The resulting mean differences and JZS Bayes factors are summarized in Table 3. The full analyses can be found in Appendix B.

As can be seen, the predictions of the RH were only partially corroborated. For those individuals classified as RH-consistent, response times did not differ between item types 1 and 2 (as predicted). However, there was a clear difference between item types 2 and 3 as well as 1 and 3—contrary to the RH-predictions. Note that, when basing analyses on the strict classification criterion (Appendix B), the evidence no longer implied robust differences between item types (although the mean difference for the comparison of item types 1 vs. 3 was comparable in size); however, in this analysis, only 10 data sets were classified as RH-consistent in the first place (see Table 2).

For datasets classified as EQW, the response time differences were in line with the predictions (which held independent of the classification criterion, see Appendix B): There was no difference between item types 1 and 2, whereas response times were longer in item type 3 as compared to both 1 and 2. Finally, as predicted, most tests favored the null-hypothesis of no response time difference for data sets classified as GUESS (the only exception is the comparison of types 1 vs. 3 when using the strict classification criterion, see Appendix B). Overall, response time patterns were very well aligned with the strategy classification in the case of EQW and satisfactorily so in the case of GUESS.

4 Discussion

Based on these findings, the conclusions of Thoma and Williams (2013) on the role of recognition in consumer choice (see also Oeusoonthornwattana & Shanks, 2010) can be further specified. Indeed, the RH does not account for the data particularly well: Although it once again turned out the superior model at least for a minority of participants in terms of accounting for choice data (for similar conclusions in probabilistic inferences, see Pachur et al., 2008), response time patterns contradicted RH predictions even for these individuals (Glöckner & Bröder, 2011; Hilbig & Pohl, 2009). Only when using a strict strategy classification criterion did response time patterns conform to the RH—while very few datasets (16%) were actually classified as RH-consistent in this case. Either way, the evidence for the RH is limited, at best. On the other hand, the very low rate of classifications for a model assuming that alternative information is weighted *more* strongly than recognition (WADD2) suggests that recognition will also rarely be overruled.

Instead, the current data suggest that recognition and further information are treated equivalently by most decision makers. Specifically, the results show that a fully specified alternative model, namely an equal weights strategy, provides a superior account of choice patterns for a modal number of data sets. Additionally, response time patterns were aligned with model predictions for these data sets, thus lending further support to the strategy classification. So, extending the conclusions of Thoma and Williams (2013), their data do favor a specific compensatory model over the RH, namely one in which recognition is weighted no more or less than the additional information (for similar conclusions, see Richter & Späth, 2006).

Nonetheless, item-level analyses also show that the strategy classification method used herein will be influenced by the specific item material. The more brands like Apple and Toshiba—which appear to be chosen irrespective of contradictory information—are included, the fewer participants will be classified as consistent with EQW (and the more with the RH). More importantly from a substantive point of view, the item-level results replicate that the RH's notion of all-or-none recognition is inappropriate (Erdfelder, Küpper-Tetzel, & Mattern, 2011; Newell & Fernandez, 2006): Familiar items come with certain knowledge which is integrated in the choice situation. By contrast, a binary all-or-none understanding of recognition could not explain why—in the face of contradictory information—brands like Apple and Toshiba were consistently chosen, whereas brands like Olympus and Shure were hardly ever chosen.

Furthermore, it should be noted that the realm of consumer choice differs in noteworthy ways from probabilis-

tic inferences for which the RH was originally proposed. In particular, when considering the preconditions for the RH proposed by Pachur et al. (2008), it is important to note that the current domain includes induced (rather than natural) cue knowledge, menu-based (rather than memory-based) information acquisition, unknown recognition validity, and additional information about unrecognized options—all of which are considered non-optimal preconditions for the RH by Pachur et al. (2008). Nevertheless, the current findings do provide insight on the role of recognition in consumer decision making—especially in providing an alternative account of how recognition information may be integrated in these decisions.

Overall, the novel findings from the current model comparison answer the call for specifying an alternative model that provides a better account of the data as compared to the RH (Brighton & Gigerenzer, 2011; Marewski et al., 2010; Pachur, 2011)—in this case, in consumer choice. Importantly, the current findings should not be over-interpreted as evidence that decision makers *used* EQW. The current methodology is not primarily designed to test any one specific process model critically, but rather to provide relative insight on which of several models provides a better account for the data. All models are necessarily abstractions and even a model that accounts for the data well need not represent the true underlying process (Roberts & Pashler, 2000). However, Thoma and Williams' (2013) data and the current analyses demonstrate that an alternative to the RH is superior in accounting for the data. In simple terms, this finding is not necessarily a strong argument *for* EQW—and indeed, a strict interpretation of EQW did not hold for each and every item—but it is stronger evidence *against* the RH as compared to the original conclusion.

References

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57–86.

Brighton, H., & Gigerenzer, G. (2011). Towards competitive instead of biased testing of heuristics: A reply to Hilbig and Richter (2011). *Topics in Cognitive Science, 3*, 197–205.

Bröder, A. (2010). Outcome-based strategy classification. In A. Glöckner & C. Witteman (Eds.), *Foundations for tracing intuition: Challenges and methods.* (pp. 61–82). New York, NY: Psychology Press.

Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta Psychologica, 121*, 275–284.

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review, 14*, 895–900.

Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making, 16*, 193–213.

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

Davis-Stober, C. P., & Brown, N. (2011). A shift in strategy or "error"? Strategy classification over multiple stochastic specifications. *Judgment and Decision Making, 6*, 800–813.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie - Journal of Psychology, 217*, 108–124.

Erdfelder, E., Küpper-Tetzel, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making, 6*, 7–22.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6*, 100–121.

Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making, 4*, 186–199.

Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making, 6*, 23–42.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75–90.

Hilbig, B. E. (2010). Reconsidering "evidence" for fast-and-frugal heuristics. *Psychonomic Bulletin & Review, 17*, 923–930.

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2011). Fluent, fast, and frugal? A formal model evaluation of the interplay between memory, fluency, and comparative judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 827–839.

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2012). A matter of time: Antecedents of one-reason decision making based on recognition. *Acta Psychologica, 141*, 9–16.

Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1296–1305.

Hilbig, B. E., & Richter, T. (2011). Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009). *Topics in Cognitive Science, 3*, 187–196.

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: extending and testing recognition-based models for multi-alternative inference. *Psychonomic Bulletin & Review, 17*, 287–309.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54.

Moshagen, M., & Hilbig, B. E. (2011). Methodological notes on model comparisons and strategy classification: A falsificationist proposition. *Judgment and Decision Making, 6*, 814–820.

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentially of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19*, 333–346.

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 923–935.

Oeusoonthornwattana, O., & Shanks, D. R. (2010). I like what I know: Is recognition a non-compensatory determiner of consumer choice? *Judgment and Decision Making, 5*, 310–325.

Pachur, T. (2011). The limited value of precise tests of the recognition heuristic. *Judgment and Decision Making, 6*, 413–422.

Pachur, T., Bröder, A., & Marewski, J. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making, 21*, 183–210.

Platzer, C., & Bröder, A. (2012). Most people do not ignore salient invalid cues in memory-based decisions. *Psychonomic Bulletin & Review, 19*, 654–661.

Pohl, R. F. (2011). On the use of recognition in inferential decision making: An overview of the debate. *Judgment and Decision Making, 6*, 423–438.

Pohl, R. F., Erdfelder, E., Hilbig, B. E., Liebke, L., & Stahlberg, D. (2013). Effort reduction after self-control depletion: The role of cognitive resources in use of simple heuristics. *Journal of Cognitive Psychology, 25*, 267–276.

Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 150–162.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367.

Roulder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Thoma, V., & Williams, A. (2013). The devil you know: The effect of brand recognition and product ratings on consumer choice. *Judgement and Decision Making, 8*, 34–44.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44*, 92–107.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53–85.

Appendix A

Model equations for use in the multiTree freeware tool (Moshagen, 2010) for each of the models considered. The first column refers to the item type (see first row of Table 1), the second column specifies the category number (i.e. which option is chosen, with 1, 3 and 5 referring to option A and 2, 4, and 6 to option B) see second row of Table 1), and the third specifies the choice probabilities. Note that the probability of choosing in line with a model's predictions is 1—*e*, as *e* denotes a strategy execution error.

RH/WADD1:	1	1	(1-e)		GUESS:	1	1	.50
	1	2	e			1	2	.50
	2	3	(1-e)			2	3	.50
	2	4	e			2	4	.50
	3	5	(1-e)			3	5	.50
	3	6	e			3	6	.50

EQW:	1	1	(1-e)		WADD2:	1	1	(1-e)
	1	2	e			1	2	e
	2	3	(1-e)			2	3	(1-e)
	2	4	e			2	4	e
	3	5	.50			3	5	e
	3	6	.50			3	6	(1-e)

Appendix B

Results of response time comparisons for all pairs of item types, conditional upon strategy classification (separately for the lenient and strict classification criterion).

Criterion	Class.	Pair	Mean diff.	s.d.	s.e.	t-value	df	p	Bayes factor
lenient	RH	neut/pos	0.004	0.092	0.021	0.20	18	0.847	5.62
lenient	RH	neg/neut	0.041	0.074	0.017	2.39	18	0.028	0.53
lenient	RH	neg/pos	0.045	0.080	0.018	2.43	18	0.026	0.49
lenient	EQW	neut/pos	0.020	0.069	0.014	1.40	23	0.175	2.56
lenient	EQW	neg/neut	0.044	0.097	0.020	2.23	23	0.036	0.72
lenient	EQW	neg/pos	0.064	0.083	0.017	3.79	23	0.001	0.03
lenient	GUESS	neut/pos	0.019	0.055	0.016	1.17	11	0.267	2.52
lenient	GUESS	neg/neut	0.030	0.105	0.030	0.98	11	0.348	3.00
lenient	GUESS	neg/pos	0.048	0.090	0.026	1.86	11	0.090	1.12
strict	RH	neut/pos	0.028	0.076	0.024	1.15	9	0.278	2.38
strict	RH	neg/neut	0.014	0.068	0.021	0.63	9	0.543	3.57
strict	RH	neg/pos	0.041	0.073	0.023	1.80	9	0.106	1.16
strict	EQW	neut/pos	0.020	0.088	0.018	1.10	23	0.284	3.61
strict	EQW	neg/neut	0.043	0.097	0.020	2.17	23	0.041	0.80
strict	EQW	neg/pos	0.063	0.090	0.018	3.40	23	0.002	0.07
strict	GUESS	neut/pos	0.011	0.058	0.015	0.76	15	0.460	4.04
strict	GUESS	neg/neut	0.040	0.098	0.024	1.65	15	0.120	1.60
strict	GUESS	neg/pos	0.051	0.083	0.021	2.48	15	0.026	0.45

Lenient refers to a maximum strategy execution error of .40, strict refers to .20 (see Table 2). Class is the classified model. S.d. and s.e. are those of the mean difference. The Bayes factor is approximated using http://pcl.missouri.edu/bf-one-sample.