A multinomial model of applying recognition to judge between multiple alternatives.

C. Philip Beaman (c.p.beaman@reading.ac.uk)
School of Psychology & Clinical Language Sciences, University of Reading
Earley Gate, Whiteknights, Reading RG6 6AL, UK

Abstract

Proponents of the "fast and frugal" approach to decision-making suggest that inferential judgments are best made on the basis of limited information. For example, if only one of two cities is recognized and the task is to judge which city has the larger population, the *recognition heuristic* states that the recognized city should be selected. In preference choices with >2 options, it is also standard to assume that a "consideration set", based upon some simple criterion, is established to reduce the options available. A multinomial processing tree model is outlined which provides the basis for estimating the extent to which recognition is used as a criterion in establishing a consideration set for inferential judgments.

Keywords: Decision-making; n-AFC; MPT models; Recognition heuristic.

Introduction

A much studied approach to the problem of determining how individuals (or groups) judge between two options is the "fast and frugal" heuristic framework advocated by Gigerenzer and colleagues (e.g., Gigerenzer, Hertwig & Pachur, 2011). This framework aims to provide both descriptive and normative accounts of how judgments are made in the real world, where time and computational resources may preclude optimizing strategies (e.g., multiple regression) where all possible sources of information are consulted and appropriately weighted combinations of these are applied to inform the judgments made.

One of the most basic principles underlying the fast and frugal approach is the *Recognition Principle*, later formalized as the Recognition Heuristic (RH; Goldstein & Gigerenzer, 2002) and implemented in ACT-R (Marewski, & Mehlhorn, 2011; Schooler & Hertwig, 2005). The RH acts as a stopping rule for more complex heuristics (e.g., Take-the-Best, Gigerenzer & Goldstein, 1996) and states simply that, "If the task is to judge which of two options scores highest on a given criterion, and only one of the two options is recognized, infer that the recognized option scores highest on the criterion". For example, the heuristic might be used to judge which of two cities has the larger population (Goldstein & Gigerenzer, 2002), which of two individuals is the wealthiest (Frosch, Beaman & McCloy, 2007) or which of two diseases occurs most frequently per year (Pachur & Hertwig, 2006).

The heuristic is applied when the criterion in question is known to correlate positively with the probability of recognition (Volz, Schooler, Schubotz, Raab Gigerenzer & von Cramon, 2006), although the reverse inference can also be made in the unlikely event that the correlation is known to be negative (Goldstein & Gigerenzer, 2002). In either case, the heuristic takes advantage of information latent in the environment to inform inference and minimize the amount of knowledge sought or retrieved from memory (or other, external sources) about the items in question prior to making a judgment.

The RH has not been universally accepted as descriptive account of human inference, however. Hilbig and colleagues, in particular, have drawn attention to the difficulty in identifying when recognition *per se* is used to inform judgment rather than knowledge which accords with recognition (Hilbig, 2010; Hilbig, Erdfelder & Pohl, 2010; Hilbig & Pohl, 2008; Hilbig, Pohl & Bröder, 2009; Hilbig & Richter, 2011). Given the positive correlation between the probability of recognition and the criterion in question, it is inevitable that many of the other cues that could in principle be consulted (even if only one of the items is recognized) will result in the same inference as recognition alone.

One way of addressing this problem is applying a multinomial processing tree (MPT) model (e.g., Hilbig et al., 2010). MPT models assume sequential, independent operations (akin to additive factors logic; Sternberg, 1998) which can be expressed as a tree structure, with alternative processes at each branch point each associated with a parameter that represents the probability of traversing that particular branch. The tree structure terminates with an observable outcome, and the models are compared to the data by estimating the best-fitting parameters and comparing the frequency counts of each outcome in experimental data to the expected outcomes given the parameters estimated (for formal reviews, see Bachelder & Rifer, 1999).

As a concrete example, consider the position in a two-alternative forced choice task (2-AFC) where the subject is asked to indicate which of the two options scores highest on the criterion of interest. Here, three trees need to be drawn. In the scenario described by the first tree, the subject recognizes neither of the options and is forced to guess between them. There are therefore only two possible outcomes, a correct guess or an incorrect guess, and hence only two branches to the tree. The probability of a correct guess (the branch leading to the frequency count of correct answers) is associated with the parameter g, and the probability of an incorrect guess is therefore given by $(1-g)$. Where guessing is truly random, $g=.5$. The scenario described by the *third* tree is similarly of little direct interest to the study of the RH. Here, the subject recognizes both options and uses knowledge to choose between them, with probability correct given by parameter b ("*knowledge validity*") and the same two-branch tree-structure as previously. This tree is potentially open to expansion, assuming that suitable candidate knowledge processes can be identified. Of more immediate interest is the *second* decision tree, which describes the scenario in which exactly one of the two options is recognized. Here, there are four possible outcomes: the recognized item might be chosen or not, and the choice might be correct or incorrect. There are also multiple ways in achieving these outcomes.

Suppose the RH is applied uncritically, then the probability that the RH produces the correct answer is given by parameter *a* ("*recognition validity*"). However, whether the RH is applied uncritically is itself governed by a parameter, *r*. Thus, the probability the RH is applied and is correct is given by *r.a* and the probability the RH is applied but results in an incorrect inference is given by *r*(1-*a*). If the RH is not applied uncritically, then the inference that the recognized item is correct might be made by the alternative route of applying knowledge. The probability that this occurs is the joint probability of the RH being valid, knowledge being valid, and knowledge being used, given by (1-*r*).*b*.*a*. The routes by which decisions might be made according to this MPT model are given exhaustively in Table 1.

Table 1:Probabilities of choosing correctly or incorrectly via different routes of three simple decision trees. The probability of arriving at the end point of any particular branch of a decision-tree is the product of the parameters associated with each branch-point.

Recognize 0: *g* (correct)	
	(1-*g*): (incorrect)
Recognize 1: *r.a* (choose recognized, correct)	
	r(1-*a*) (choose recognized, incorrect)
	(1-*r*)*b*.*a* (choose recognized, correct)
	(1-*r*)*b*(1-*a*) (choose unrecognized, correct)
	(1-*r*)(1-*b*)*a* (choose unrecognized, incorrect)
	(1-*r*)(1-*b*)(1-*a*) (choose recognized, incorrect)
Recognize 2: *b* (correct)	
	(1-*b*) (incorrect)

From this table, it is clear that MPT models are well-suited to answering the question of the extent to which the RH is employed in any given 2-AFC inference task. Estimation the best-fitting set of parameters from the data provides the most likely rate of application of the RH in terms of the *r* parameter. This can be done across the sample of all subjects within the experiment or, where a sufficiently large data-set is available for estimations to be made on an individual basis, it can be done per subject. This may be useful if, for example, *r* and *a* are not independent for some subjects (see, e.g., Beaman et al., 2010). However, because of the limited data available, in the study reported below only estimation across the full dataset was attempted.

Multi-Alternative Inferences

Not all inferences of the kind described above involve choosing between only two alternatives. Frosch et al. (2007) presented 3- or 4-alternatives to choose from in a wealth judgment task. Marweski et al. (2010) likewise examined the incidence of RH use when judging likelihood of electoral success amongst named politicians. Theoretical analyses of the success of the RH as a normative theory under these circumstances are also provided by McCloy, Beaman & Smith (2008). Both *a priori* theoretical analyses and empirical studies rely upon the presumption that there are independent processing stages in which *N* alternative options are reduced to a smaller number from which to choose, but the decision-making processes employed to reduce the number of alternative options may differ from those employed to choose from amongst those options. This situation parallels one in the literature on consumer choice, in which it is assumed that a "consideration set" is initially formed from within all the options available (Howard, 1963; Wright & Barbour, 1977). Indeed, in some circumstances, criterion-inference judgments are equivalent to preference choices (e.g., if subjects are asked to indicate which politician they believe will win a given election, as in Marewski et al.'s (2010) data, they are *de facto* being asked which candidate they believe the public as a whole will prefer to win the election). These circumstances also lend themselves to modeling by means of MPTs. One difference between criterion-inference and preference-choice ,ay lie in the way in which criteria are applied for preference-choice. However, this does not preclude the possibility that consideration sets may be invoked in multi-alternative inference, even if the nature of the criteria applied may need not be identical across the two situations.

The structure of 3-AFC judgments.

To model such multi-alternative inferences using MPTs, an appropriate data-set is required and tree structures representing the structure of the model must be constructed. Here, the example of 3-AFC inference is explored using data from Frosch et al. (2007) in a study in which subjects were asked to indicate which of three names taken from the *Sunday Times* rich list (an annual compilation of the richest individuals in the UK) was the wealthiest. 27 such names were chosen and, from within this group, were randomly 22 lists of three names each were produced by random selection. 26 subjects were then each given the same series of three names to choose from per trial and were asked to indicate which name they thought was the wealthiest. They were then re-presented with all the names and asked to indicate which names they had known prior to the beginning of the experiment. As with 2-AFC, the trees describing scenarios in which either all the options or none of the options are recognized are relatively uninteresting, being governed purely by single guessing or knowledge parameters (*g* and *b*, respectively).

For 3-AFC scenarios in which one item is recognized, a plausible tree structure is given in Figure 1.The tree structure when one item from three is recognized differs from the corresponding structure when one item is recognized in a 2-AFC task. As shown in Figure 1, there is an extra step with 3-AFC when the RH is not applied, knowledge is valid (i.e., will provide the correct inference) but the RH is not valid (will provide an incorrect inference). The reasons for this can readily be discerned as follows: If knowledge is inaccurate and is used in preference to the RH, then it will always lead to a false choice, either of the recognized item (if recognition is also an invalid cue) or one of the unrecognized items (if recognition is a valid cue), as in the 2-AFC situation. However, if knowledge is valid then this knowledge must be either that the one recognized item is high on the criteria or that the recognized item is low on the criteria. If the former, the recognized item will be chosen, if the latter then it will be excluded from further analysis. If both knowledge and recognition provide valid cues, then the recognized item must be the correct item and it will be chosen. If, however, recognition is not a valid cue then the correct item must be one of the two unrecognized options: hence there is an extra guessing stage to choose between these two. In terms of the equations presented in Table 1, the conditional probability of guessing correctly when following this particular branch of the tree can be expressed as, $(1\text{-}r_1).b_1.(1\text{-}a_1).g_{1,}$

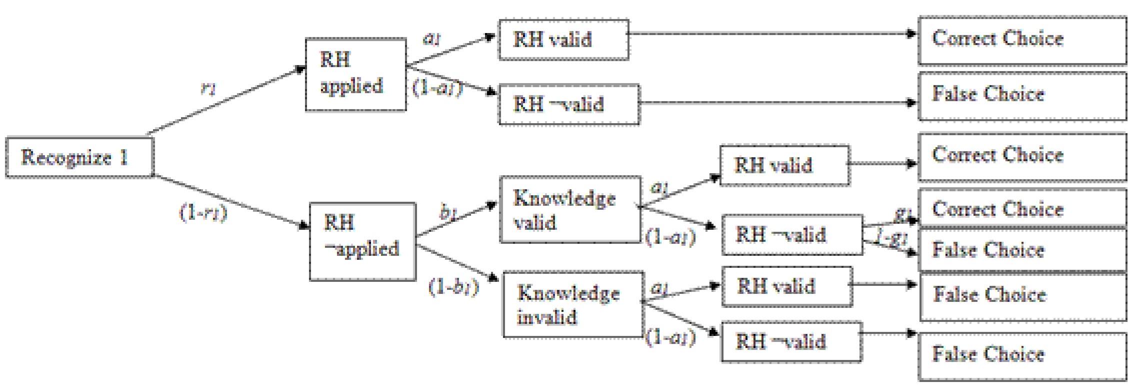

Figure 1. A 3-AFC decision tree indicating possible processing paths leading to correct or incorrect criterion inferences when exactly one item from three is recognized.

A similar logic can be applied to construct the tree structure for scenarios where two of the three options were recognized. The decision tree for the scenario when two items (out of three) are recognized is more complex than when only one item is recognized (Figure 2). Here, the *RH inclusion rule* is explicitly distinguished from the RH itself. The RH inclusion rule operates to winnow down the number of options under active consideration to a consideration set defined in terms of whether the items were recognized or not. Thus the RH inclusion rule operates along the recognition principle, as does the RH, but unlike the RH is only the first step in establishing – by a longer chain of inference – which of the items scores most highly on the criterion. It is an empirical question whether the RH inclusion rule and the RH are applied at equivalent rates amongst the same group of subjects attempting the same criterion judgment task.

Figures 1 and 2 show how inferences may be made on the basis of recognition, knowledge, or recognition plus knowledge when three items are presented but recognition is incomplete. Under these circumstances, of course, the chosen item may also be either recognized or unrecognized and tracing the relevant branch of the decision tree gives this information also (e.g., if valid knowledge is used and recognition is also a valid cue then the correct choice made must be the choice of the recognized option).

When the inclusion rule is applied, recognition may be a valid or an invalid cue. If it is invalid, then straightforwardly, the choice of item must be incorrect as only the two recognized items are under consideration. If, however, recognition is a valid cue then correct inference is still not guaranteed as a choice must be made between the two recognized items. If the inclusion rule is not applied, then the knowledge about the two recognized items consulted instead may also be either valid or invalid. If it is valid and recognition is also a valid cue, that is to say the correct item is one of the two recognized items, then a correct inference is made because valid knowledge of the two items is sufficient to choose between them. If the inclusion rule is invalid, then valid knowledge must be of a different type – knowledge that neither of the recognized items is correct. In this case too, however, correct inference is guaranteed. Following the same lines of argument, if knowledge is invalid then an incorrect choice is made regardless of whether recognition is valid; that is, regardless of whether the invalid choice is of a recognized or unrecognized item.

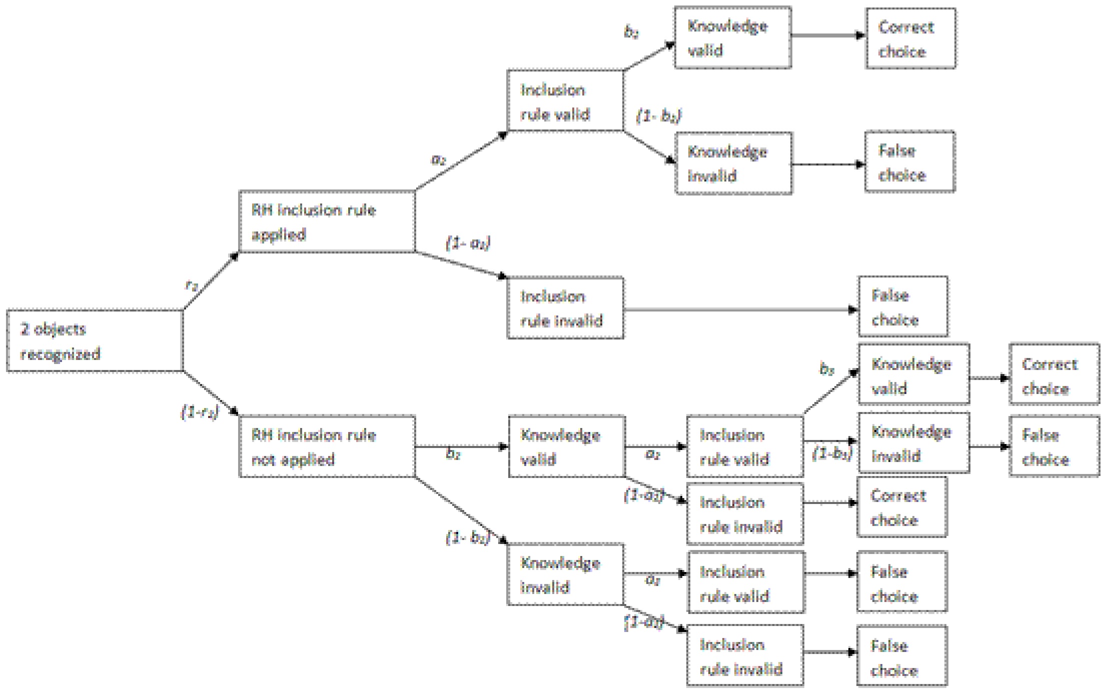

Figure 2. A 3-AFC decision tree indicating possible processing paths leading to correct or incorrect criterion inferences when exactly two items from three are recognized

Comparing the Model to "Rich List" Data

The best-fitting parameters for the model were estimated from Frosch et al.'s (2007) 3-AFC study using MultiTree software (Moshagen, 2010). "Guessing" parameters for situations in which neither knowledge nor recognition could be employed were set to produce chance-level accuracy (i.e., $g = .33$, $g_1 = 0.5$).

An initial model assumed a single value for knowledge validity to determine – based on knowledge – whether one of two recognized items was correct and also *which* of these two recognized items was the correct option (i.e., on Figure 2, $b_3 = 1.0$). For this model, recognition and knowledge validities were separately estimated for each recognition scenario (recognize 0, 1, 2 or 3) as there are reasons to suppose that knowledge validity should vary as a function of the number of items recognized (e.g., Smith, Beaman & McCloy, 2011). Estimating all parameters other than the "guessing" parameters, g and g_1, a goodness-of-fit test shows that the best-fitting version of this model nevertheless produced expected results which differed significantly from those observed, G^2=11.01, df=3, p = .01.

A second model, assuming a decision-point (with validity b_2) that one of the recognized items – based on knowledge – was the correct item and a further decision-point (again, based on knowledge but with a different validity, b_3) produced expected results that did not differ significantly from those observed, G^2=1.05, df=2, p=.59.

Knowledge validity when all three items were recognized was estimated as 0.47 and when one item of the three was recognized knowledge validity was 0.56. Knowledge validity for correctly realizing –based on knowledge – that one of the two recognized items was the correct choice was high, at 0.91 but choosing between these two items once this decision was made was difficult, with validity of choices estimated at 0.30.

Interestingly, for this model, the probabilities r_1 and r_2 that the RH and the RH-inclusion rule were employed (Figures 1 and 2) were estimated as .64 and .63 respectively, and the validities of these two rules were also estimated as being very similar, a_1 (RH validity) = .72 and a_2 (inclusion rule validity) = .73. Constraining these two rules and their validities to take the same values resulted in a model where a=.73 and r=.63 which did not differ significantly from the baseline model, ΔG^2 = .19, df = 2, p = .91, nor from the data, G^2=1.25, df=4, p=.88. For the 3-AFC situation, therefore, it appears that there is no discernible difference between using the RH to guide choice when only one item is recognized and using the RH-inclusion rule to form a consideration set which is subsequently informed by knowledge. Full parameter values for this version of the model are given in the appendix, a comparison of the model to the data in terms of observed and expected frequencies with which recognized and unrecognized items were chosen either correctly or incorrectly is given in Table 2. Although parameter overfitting is a legitimate concern in any modelling domain, on which may be fuelled by observation of the close fits between expected and obtained results here, it is worth noting that – like standard hypothesis testing – a failure to provide a good fit given certain assumptions (e.g. that r=1.0) may be as informative as obtaining a good fit, and it is on this basis that the G^2 goodness-of-fit statistic operates.

Table 2: The frequency of responses tabulated according to whether they were correct or not, and whether the option chosen was recognized or not. Observed responses are taken directly from the data of Frosch et al. (2007), expected responses are the frequencies of responses within those particular categories given a model with the tree structures shown in Figures 1 and 2 and the parameters given in the appendix.

	Frequency of Responses	
	Observed	*Expected*
Recognize 0:		
Correct	40	35
Incorrect	65	70
Recognize 1:		
Correct (recog)	83	84.40
Incorrect (recog)	25	23.69
Correct (¬recog)	10	9.87
Incorrect (¬recog)	16	16.19
Recognize 2:		
Correct (recog)	79	77.81
Incorrect (recog)	62	62.79
Correct (¬recog)	14	14.32
Incorrect (¬recog)	4	4
Recognize 3:		
Correct	81	81
Incorrect	93	93

Finally, models in which the use of the RH was set at the level of RH use indicated by the data, if choice of the recognized item was taken as an index of RH-use (i.e., r_1=.78), differed significantly from the best-fitting model, ΔG^2=5.15, df=1, p=.02, but not from the data, G^2=6.2, df=3, p=.1. However, if the same exercise is repeated for the RH-inclusion rule, the resulting model differs both from the best-fitting model, ΔG^2=13.59, df = 1, p=.0002, and from the data, G^2=14.65, df=3, p=.002. Using RH-adherence rate as an index of the RH-inclusion rule is therefore likely to be misleading.

Discussion

The modeling results reported here indicate the viability of applying the recognition principle as a means of forming a "consideration set" in criterion judgment as well as preference judgment, when there are more than two alternatives to consider. The particular tree structures proposed for an independent series of decisions can also be fit to the data with estimated parameters providing expected results that do not differ significantly from those observed in the data (Table 2). Although the particular data-set modeled here (relative wealth judgments) clearly require a criterion judgment, the distinction between criterion and preference judgments is less distinct with other stimuli. For example, Marewski et al. (2010) apply the RH to understanding judgments about the relative success of candidates in German elections. Although this is clearly a criterion judgment ("who is most likely to win?"), it can also be framed as a preference judgment ("who is the best candidate?") whilst still potentially meeting the pre-requisites for applying the recognition heuristic or recognition principle ("infer that the recognized option scores highest on the criterion"). Elucidating the means by which multi-alternative inferences may be made, even if the expansion is the relatively modest one from 2-AFC to 3-AFC, helps indicate where there are possible similarities between preference choice (where multiple alternatives are frequently presented) and criterion inference (where, typically, paired choices are presented, as in the "city size" task popularized by Gigerenzer and Goldstein (1996).

One interesting feature that became apparent when modeling these data was the need for multiple different parameters to represent knowledge validity, if not recognition validity. This both supports Smith et al.'s (2011) arguments that knowledge validity should vary as a function of the number of items recognized (see also Beaman et al., 2010) and highlights the need for different types of knowledge to be consulted to choose that the correct item is one of the recognized set (e.g., b_2) and which member of the recognized set the correct item may be (e.g., b_3; $b_2 \neq b_3$).

The model also indicates that, at least for the data-set considered here, there may be little or no difference between the application and validity of a "recognition inclusion rule" and the recognition heuristic *per se*. This is something of a surprise as, *a priori*, one would assume that a recognition inclusion rule might be employed, as a cognitive short-cut, even more frequently than the RH itself. However, it may simply be that the 3-AFC was too similar to 2-AFC for any differences to become apparent. The data also show that recognition adherence rate is not a good measure of use of the recognition principle when more than one item is recognized and a recognized item is chosen. The resulting model, where r_2 is set at the adherence rate, differs significantly from both the baseline, best-fitting model and the data. Setting r_1, the incidence of employing the RH, also produces a model that differs significantly from the baseline model although, in this case, both models fit the data.

Finally, it is worth noting some limitations of the current modelling approach. Like all formal approaches, it relies upon a number of background assumptions. The need for multiple parameters and the relative independence of certain parameters have already been alluded to, but of equal concern is the lack of an underlying process model. The framework as it currently exists blurs the boundaries between the existence of cognitive processes and their efficacy, and the challenge is to provide a process model which both draws clear distinctions between the two and shows how, and when, knowledge is consulted.

Acknowledgments

The multinomial modeling research reported here was supported by ESRC grant RES-062-23-1752. The experimental data to which the model was fit was funded by the Leverhulme Trust, grant no. F/00 239/U. I am grateful to Caren A. Frosch for collecting and analyzing these data and to Philip T. Smith for many useful discussions.

References

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

Beaman, C. P., Smith, P. T., Frosch, C. A., & McCloy, R. (2010). Less-is-more effects without the recognition heuristic**.** *Judgment & Decision-Making, 5*, 258-271.

Frosch, C., Beaman, C. P., & McCloy, R. (2007). A little learning is a dangerous thing: An experimental demonstration of ignorance-driven inference. *Quarterly Journal of Experimental Psychology, 60*, 1329-1336

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

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford: OUP.

Goldstein, D., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 103*, 650-669.

Hilbig, B. E. (2010). Precise models deserve precise measures: a methodological dissection. *Judgment and Decision Making, 5*, 272-284

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision-making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory & Cognition 36*, 123-134

Hilbig, B. E. & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology, 55*, 394-401

Hilbig, B. E., Pohl, R. F. & Bröder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making, 22*, 510-522

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

Howard, J. A. (1963). *Consumer Behaviour: Application of theory*. New York: McGraw-Hill.

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

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.

Marewski, J. N., & Mehlhorn, K. (2011). Using the ACT-R architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making, 6*, 439-519

McCloy, R., Beaman, C. P., & Smith, P. T. (2008). The relative success of recognition-based inference in multi-choice decisions. *Cognitive Science, 32*, 1037-1048

Pachur, T. & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*, 983-1002.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*, 610-628.

Smith, P. T., Beaman, C. P., & McCloy, R. (2011). *The recognition heuristic and knowledge use*. Poster presented to Subjective Probability, Utility & Decision-Making (SPUDM) 23, Kingston-upon-Thames, London.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In: D. Scarborough & S, Sternberg (Ed.s) *An invitation to cognitive science, volume 4: Methods, models and conceptual issues*. Cambridge, Ma.: MIT Press.

Wright, P. & Barbour, T. (1977). Phased decision strategies: Sequels to initial screening. In: M. Starr & M. Zeleny (Ed.s) *Multiple criteria decision making*. Amsterdam: North Holland Publishing.

Appendix

Parameter values for the version of the 3-AFC model fit to the data in Table 2. Standard errors for all of the estimates are given in parentheses.

Free Parameters:	
Probability applying recognition principle	
r_1 (RH) = r_2 (inclusion rule)	.63 (.05)
Recognition validity for RH and recognition inclusion rule	
a_1=a_2	.73 (.03)
Knowledge validity	
b_1 (1 item recognized)	.56 (.11)
b_2 (2 items recognized)	.91 (.05)
b_3 (choice between 2 recognized items)	.47 (.04)
b_4 (3 items recognized)	.30 (.17)
Fixed Parameters:	
Probability correct guess	
g	.33
g_1	.5