Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method

Andreas Glöckner*

Max Planck Institute for Research on Collective Goods

Abstract

One of the core challenges of decision research is to identify individuals' decision strategies without influencing decision behavior by the method used. Bröder and Schiffer (2003) suggested a method to classify decision strategies based on a maximum likelihood estimation, comparing the probability of individuals' choices given the application of a certain strategy and a constant error rate. Although this method was shown to be unbiased and practically useful, it obviously does not allow differentiating between models that make the same predictions concerning choices but different predictions for the underlying process, which is often the case when comparing complex to simple models or when comparing intuitive and deliberate strategies. An extended method is suggested that additionally includes decision times and confidence judgments in a simultaneous Multiple-Measure Maximum Likelihood estimation. In simulations, it is shown that the method is unbiased and sensitive to differentiate between strategies if the effects on times and confidence are sufficiently large.

Keywords: strategy classification, judgment, decision making, maximum likelihood estimation, intuition.

1 Methods for strategy classification

In different situations, people might use different strategies to decide. These strategies might sometimes be completely based on conscious processes, such as comparing the available options on the most important attribute and choosing the option that is better on this attribute (e.g., Beach & Mitchell, 1978; Fishburn, 1974; Payne, Bettman, & Johnson, 1988), or people might rely more or less on automatic processes that integrate information unconsciously (e.g., Busemeyer & Townsend, 1993; Dougherty, Gettys, & Ogden, 1999; Glöckner & Betsch, 2008b). Decision researchers often are interested in the question which strategy was (more likely) used by each person. Several methods have been suggested to identify decision strategies. The three predominant approaches are structural modeling, process tracing, and comparative model fitting (for overviews see Bröder & Schiffer, 2003a; Glöckner & Witteman, in press; Harte & Koele, 2001).

1.1 Structural modeling

Structural modeling uses a multiple regression approach to identify how cues or attributes are utilized in making judgments (Brehmer, 1994; Doherty & Brehmer, 1997; Doherty & Kurz, 1996). Specifically, a set of judgments (*criterion*) is predicted by cue values (*predictors*). Regression weights can be interpreted as indicators for individuals' usage of cues in their judgments. Structural modeling usually does not aim to analyze processes (as in the paramorphic approach; see Hoffman, 1960) but input-output relations between cues and judgments only (i.e., as-if models; but see Bröder, 2000, Exp. 1). Although the method was tremendously useful in showing that people integrate information in a weighted compensatory manner when making more or less intuitive judgments (Doherty & Brehmer, 1997; Hammond, Hamm, Grassia, & Pearson, 1987), its focus on outcomes limits its applicability for differentiating among process models.

1.2 Process tracing

Process tracing methods record and analyze parameters of information search before judgments or decisions and aim to infer decision strategies from the amount, distribution and order of information search. For instance, information boards are often used in which information is provided behind hidden information cards, which are

opened on request or by mouse-click (e.g., Payne et al., 1988; Rieskamp & Hoffrage, 1999).

This method allows differentiating between decision strategies because some of them differ in their predictions concerning information search. A simple take-the-best strategy (TTB, Gigerenzer & Goldstein, 1996; cf. Fishburn, 1974), for example, assumes that persons first look up the predictions of the most predictive (*valid*) cue for all options. The option with the best cue value is selected. If options are tied, the second most valid cue is considered, and so on. In contrast, according to an equal weight strategy (EQW, Payne et al., 1988), individuals look up all cue information for the first option and sum them up. Then they do the same for the second option and so on and select the option with the highest sum of cue values. Hence, a cue-wise information search (i.e., search along cues) and a strong focus on the most valid cue are used as indicators for the usage of TTB (or similar non-compensatory strategies) and an option-wise information search and an equal inspection of all cues indicate the usage of EQW (or other compensatory strategies).

Despite being highly useful for investigating deliberate strategies, standard process-tracing methods such as Mouselab (Payne et al., 1988) influence strategy selection and hinder people from applying intuitive-automatic decision strategies (Glöckner & Betsch, 2008c; see also Norman & Schulte-Mecklenbeck, in press). One reason for this is that classic process tracing methods induce a serial information search and prevent quick comparisons between options and the formation of holistic impressions. One might argue that intuitive-automatic processes cannot be captured by the analysis of information search at all. This conclusion, however, seems to be too strong considering the successful use of less intrusive methods such as eye-tracking technology to investigate intuitive processes (e.g., Glöckner & Herbold, 2008). Eye-tracking methods even provide further dependent measures (e.g., eye-fixation duration and physiological arousal; Horstmann, Ahlgrimm, & Glöckner, under review), which could be included in strategy classification.

1.3 Comparative model fitting

The more recently-developed comparative-model-fitting approach uses a maximum likelihood method to compare choices with the predictions of a set of decision strategies (Bröder & Schiffer, 2003a; Bröder, in press; Wasserman, 2000). For instance, assume one observes choices in 10 decisions from which 6 are in line with the predictions of TTB and 8 are in line with the prediction of EQW. An obvious scheme would be to classify persons according to the amount of strategy compatible choices — with 8/10 choices in line with EQW, and only 6/10 in line with TTB,

we would classify this person as consistent with EQW. However, this simple counting method leads to biased results if strategies predict random choices for some of the tasks (Bröder & Schiffer, 2003a; Bröder, in press).

Maximum likelihood estimation provides a more elegant means of performing strategy classification that is not prone to this source of bias. The basic idea is simple: comparative model fitting determines the strategy that would most likely have produced the observed choice pattern under the assumption of a constant error rate in applying the strategy. For the example above, the best estimates for the error rates in strategy application would be .40 (i.e., 4/10 "errors" in applying a TTB strategy) and .20, respectively. According to the binomial equation (see also equation 1, below), the likelihood that exactly the observed number of strategy conforming choices (6 out of 10 correct under the assumption of an error rate of .40) was produced by TTB is .25 whereas the respective likelihood for EQW is .30. Hence, it is more likely that choices were produced by application of EQW than by TTB.

In contrast to classic process-tracing methods, the comparative model fitting approach avoids influencing decision behavior by the measuring method and nevertheless allows process models to be tested. However, the method is applicable only when strategies make different choice predictions. If strategies make the same predictions for choices, the likelihoods for the strategies will obviously always be equal.

Unfortunately, strategies often make exactly the same choice predictions. This is due to the fact that the principally investigated decision strategies (e.g., TTB and EQW) are special cases of a weighted additive strategy (WADD). According to a WADD strategy, for each option the cue information is weighted by its importance (or validity) and added up. The option with the highest weighted sum is chosen (Payne et al., 1988). Although this strategy sounds quite different from TTB and EQW, it can be easily shown that WADD predicts the same choices as TTB. This is always the case if the validity of each cue is higher than the sum of the validity of all less valid cues (Bröder, 2000; Lee & Cummins, 2004). Similarly, WADD predicts the same choices as EQW if the validity of all cues is similar or equal. Hence, in a strict sense, classification for EQW and TTB based on choices only never rules out that a more complex WADD strategy was used. A person could have used WADD with a specific cue weighting scheme instead. Therefore, in all studies relying on the choice-based strategy classification method (e.g., Bröder & Schiffer, 2003b; Bröder & Gaissmaier, 2007) the estimated proportions of TTB and EQW users are upper limits for the usage of these simple strategies whereas the usage of WADD is likely to be underestimated.

The problem of similar choice predictions becomes even more severe if one considers that people might also use intuitive decision strategies. Intuitive decision strategies often also predict choices that follow weighted additive information integration without the assumption that individuals calculate weighted sums (Busemeyer & Johnson, 2004; Glöckner & Betsch, 2008b; Hammond, Hamm, Grassia, & Pearson, 1987). Therefore, based on the analysis of choices only, they cannot be distinguished from WADD nor, in a strict sense, from TTB and EQW either.

In this article I aim to show that the problem can be solved in that the method for comparative model fitting based on maximum likelihood estimation of choices suggested by Bröder and Schiffer (2003a) is extended by including additional dependent measures such as decision time and confidence (for earlier approaches, see Bergert & Nosofsky, 2007; Glöckner, 2006). A Multiple-Measure Maximum Likelihood (MM-ML) strategy classification method is suggested that allows identifying decision strategies even if they make the same choice predictions and different predictions concerning only one of the other dependent variables (i.e., decision time, confidence). Further advantages of the inclusion of additional dependent measures will be discussed.

2 Examples for strategy classification

To apply a strategy classification method, it is necessary to select a set of strategies that allows for deriving predictions concerning choices, decision time, and confidence. Furthermore, item types have to be selected for which the strategies make different predictions on as many dependent variables as possible. These types have to be repeatedly presented (e.g., 10 times; Bröder & Schiffer, 2003a). In this analysis I focus on choices in probabilistic inference tasks, in which persons select the better of two goods based on recommendations of four advisors (cues) with different reliability of recommendations (i.e., cue validity). The considered strategies are WADD, TTB, EQW, a random choice strategy (RAND), and an intuitive parallel constraint satisfaction strategy (PCS; Glöckner & Betsch, 2008b). The choice predictions of PCS and WADD are essentially equal (considering different cue-validity transformation functions) and hence the strategies cannot be differentiated based on choices only. The steps to derive the strategies' predictions concerning choices, decision times and confidence are explained in detail elsewhere (Glöckner, in press) and the most important aspects will be summarized below. The resulting predictions for six types of items are shown in Table 1. It is assumed that an experiment was conducted in which

each item type was presented 10 times (in individually randomized order), and choices, decision times and confidence were recorded.

2.1 Specification of WADD

In order to derive predictions from WADD, it has to be determined how cue validities are used in calculating weighted sums. In this paper it is assumed that persons correct their weights for the fact that binary cues with a validity of .50 have no predictive power ($w = v$ - .50; cf. Glöckner & Betsch, 2008c). Although sometimes stated otherwise, choice predictions (as well as time and confidence predictions) of WADD are *not* invariant to this transformation. In the following I use the label WADD$_{corrected}$ when referring to the predictions of such a WADD strategy with corrected weights.

2.2 Decision Time Predictions

Decision-time predictions for the deliberate strategies TTB, EQW, and WADD are determined according to the number of elementary information processes necessary to apply the strategy (Payne et al., 1988). For instance, for item types 1 to 5 (see Table 1), according to TTB, only one cue has to be considered, whereas for item 6, two cues have to be considered, which necessitates applying more elementary information processes. For statistical reasons, decision time predictions are transformed to contrast weights which add up to zero and have a range of 1. For WADD and EQW, no differences in decision times are predicted and all contrast weights are set to zero. For PCS decision time predictions were derived from a simulation of the underlying network model (i.e., based on the iteration the PCS algorithm needs to find a consistent solution; Glöckner & Betsch, 2008b).

2.3 Confidence Predictions

Confidence predictions of TTB were derived based on the validity of the differentiating cue (Gigerenzer, Hoffrage, & Kleinbölting, 1991). For WADD and EQW the difference between the weighted (unweighted) sums of cue values for the two options was calculated and used as prediction for confidence. For PCS the predictions were derived from model simulations (i.e., based on the difference between the activation of the options after the consistent solution was found).

Table 1: Item types and predictions of strategies.

	Types of decision tasks					
	1	2	3	4	5	6
	A B	A B	A B	A B	A B	A B
Cue 1 ($v = .80$)	+ −	+ −	+ −	+ −	+ −	− −
Cue 2 ($v = .70$)	+ −	+ −	− +	− −	− +	− −
Cue 3 ($v = .60$)	+ −	− +	− +	− −	+ −	+ −
Cue 4 ($v = .55$)	− +	− +	− +	− +	− +	− +
	Choice Predictions					
TTB	A	A	A	A	A	A
EQW	A	A:B	B	A:B	A:B	A:B
WADD$_{corrected}$	A	A	B	A	A	A
PCS	A	A	B	A	A	A
	Time Predictions (contrasts)					
TTB	−0.167	−0.167	−0.167	−0.167	−0.167	0.833
EQW	0	0	0	0	0	0
WADD$_{corrected}$	0	0	0	0	0	0
PCS	−0.400	−0.310	0.600	−0.120	0.110	0.130
	Confidence Predictions (contrasts)					
TTB	0.167	0.167	0.167	0.167	0.167	−0.833
EQW	0.667	−0.330	0.667	−0.330	−0.330	−0.330
WADD$_{corrected}$	0.630	0.230	−0.370	0.030	−0.170	−0.370
PCS	0.620	0.280	−0.320	−0.010	−0.190	−0.380

Note. Items types and predictions of decision strategies. In the upper part of the table, the item types are presented. The cue validities v are provided beside each cue. Below the predictions concerning choices are shown. A and B stand for the predicted option. "A:B" indicates random choices between A and B. The lower part of the table shows predictions for decision times and confidences expressed in contrast weights that add up to zero and have a range of 1. Contrast values represent relative weights comparing different cue patterns for one strategy.

3   The maximum likelihood strategy classification method for choices

The maximum likelihood strategy classification method for choices calculates the conditional likelihood of an observed set of choices for different types of tasks j given the application of a certain decision strategy k and a constant error rate ε_k. The likelihood values of the different strategies are compared and individuals are classified as users of the strategy that most likely produced the observed choices. For each of the choices and each strategy, it is determined if the choice was in line with the prediction of the strategy or not. Let n_j be the number tasks of

type j that are presented and let n_{jk} be the number of correct predictions of strategy k. The likelihood of observing a certain number of correct predictions n_{jk} given a constant error rate follows a binomial distribution. Hence, the likelihood of observing a set of choices given a strategy k and a constant error rate ε_k can be calculated by:

$$L_{k(C)} = p(n_{jk}|k, \epsilon_k) = \prod_{j=1}^{J} \binom{n_j}{n_{jk}} (1-\epsilon_k)^{n_{jk}} \epsilon_k^{(n_j-n_{jk})}.$$

(1)

The single free parameter ε_k can be estimated using standard statistical software packages such as STATA or, in

this simple case, by:

$$\hat{\epsilon}_k = \left[\sum_{j=1}^{J} (n_j - n_{jk}) \right] \div \left[\sum_{j=1}^{J} n_j \right]. \qquad (2)$$

Individuals are classified as users of the strategy with the highest likelihood value $L_{k(C)}$. If a strategy does not differentiate between options for a specific type of items, individuals are assumed to guess and ε_k is assumed to be .5 for this type. Bröder and Schiffer (2003a) showed in simulations that up to an error rate of 25% the method differentiates well between strategies which make different predictions concerning choices (i.e., classification error below 20%).[1] In decision research, the method has been successfully applied to judgments and choices based on probabilistic inference (e.g., Bröder, 2003; for an overview see Bröder, in press; Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003b, 2006; Glöckner, 2006, 2007) and decisions under risk (Glöckner & Betsch, 2008a).

An earlier publication (Glöckner, 2006) highlighted the limitations of this method and made a first attempt to use decision times in individual strategy classification. To differentiate between intuitive and deliberate decision strategies with equal choice predictions, paired *t*-tests were used to compare individuals' decision times in choices for different item types, for which one strategy predicts no difference and the other does. A similar method was used in a recent work by Bergert and Nosofsky (2007). This method can be criticized in different respects: (a) it does not take into account the total fit of decision times to the total set of predictions of the strategies but is based on pair-wise comparisons of two types of items only, (b) it gives a certain strategy the advantage of the null hypothesis without controlling for the beta-error,[2] and (c) the results of the choice-based strategy classification (i.e., $L_{k(C)}$) and the *t*-test(s) (i.e., *t* and *p* value) for choices cannot easily be integrated into one single measure of fit for the strategy. While the first two problems might be circumvented using correlation measures and estimating the beta-error based on the expected effect size and number of observations, the third problem is harder to tackle (Glöckner, in press). The Multiple-Measure Maximum Likelihood strategy classification method which is introduced next solves the first and the last and reduces the second problem by using one

---

[1]Note, however, that the exact estimations of classification errors depend crucially on the number of item types and items per type (n_j).

[2]The strategy that predicts no difference in decision time is often given the advantage of being the null-hypothesis. Only if a significant difference in the direction predicted by the other strategy is found the null-hypothesis is rejected and a classification for the alternative strategy is done. As discussed elsewhere (Glöckner, in press), with small n this leads to over-classification for strategies that predict no differences because the beta-error is bigger than conventional alpha levels.

single maximum likelihood measure for choices, decision times and confidence.

4 Multiple-Measure Maximum Likelihood (MM-ML) strategy classification

Maximum likelihood estimation is, of course, not limited to dichotomous outcomes (i.e., choices) but can also be applied to continuous variables such as decision times. However, estimation of the likelihood of a set of observations necessitates assumptions about the distribution underlying the data generation process for the variable. One standard assumption is that log-transformed decision times are normally distributed (Bergert & Nosofsky, 2007, Appendix C). Under this assumption, the likelihood value of observing a log-transformed decision time x given $N[\mu, \sigma]$ can be derived from the density function of the normal distribution:

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (3)$$

and for a set of i independent observations $\vec{x}$ drawn from the same distribution by:

$$L_{k(T)} = p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \qquad (4)$$

The density function of the normal distribution (equation 3) contains two parameters. The mean is represented by μ, the standard deviation is indicated by σ (π and e are of course constants). The variable x indicates the value for which the likelihood value should be determined. According to the properties of a normal distribution, the likelihood value of x decreases with increasing distance from μ (because the exponent of e becomes a higher negative number) and it also decreases with decreasing σ. The total likelihood of events is the product of the single likelihoods of these events. Therefore in equation 4 the total likelihood for all observed decision times results from multiplying the likelihood of all single events (as indicated by the product sign).

Under the assumption that choices and decision times are independent (for a more detailed discussion of the issue of independence see below), the likelihood of observing a set of choices and decision times given the application of a strategy k, a constant error rate for choices ε_k, and decision times that are drawn from a unique normal

distribution $N[\mu, \sigma]$ is:

$$L_k = L_{k(T)}L_{k(C)} = p(n_{jk}, \vec{x}|k, \epsilon_k, \mu, \sigma) = \prod_{j=1}^{J} \binom{n_j}{n_{jk}} (1 - \epsilon_k)^{n_{jk}} \epsilon_k^{(n_j - n_{jk})} \times \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}. \quad (5)$$

Equation 5 should obviously be applied only for decision strategies such as *WADD* and *EQW*, which predict equal decision times for all considered types of tasks. Strategies *TTB* and *PCS* make (interval-scaled) predictions. Let us denote these predictions t_i and assume that they are scaled as contrast weights which add up to 0 and have a range of 1. Let us further assume that decision times for the item i are drawn from different normal distributions with means

$$\mu_i = \mu + t_i R, \quad (6)$$

in which R represents a (non-negative and to be estimated) scaling parameter. The likelihood value for observing a set of choices and decision times drawn from different normal distributions (with equal σ)[3] can then be calculated by inserting equation 6 in equation 5:

$$L_k = p(n_{jk}, \vec{x}|k, \epsilon_k, \mu, \sigma, R) = \prod_{j=1}^{J} \binom{n_j}{n_{jk}} (1 - \epsilon_k)^{n_{jk}} \epsilon_k^{(n_j - n_{jk})} \times \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - (\mu + t_i R))^2}{2\sigma^2}}. \quad (7)$$

Furthermore, assuming that confidence judgments are independent of choices and decision times and normally distributed, confidence estimation can be added to equation 7 in the same manner as decision time.[4] From extending equation 7 and adding subscript $_T$ and $_C$ for parameters referring to decision time and confidence, respectively, results in equation 8:

$$L_{total} = p(n_{jk}, \vec{x}_T, \vec{x}_C|k, \epsilon_k, \mu_T, \sigma_T, R_T, \mu_C, \sigma_C, R_C) =$$

$$\prod_{j=1}^{J} \binom{n_j}{n_{jk}} (1 - \epsilon_k)^{n_{jk}} \epsilon_k^{(n_j - n_{jk})} \times$$

$$\prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma_T^2}} e^{-\frac{(x_{T_i} - (\mu_T + t_{T_i} R_T))^2}{2\sigma_T^2}} \times$$

$$\prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma_C^2}} e^{-\frac{(x_{C_i} - (\mu_C + t_{C_i} R_C))^2}{2\sigma_C^2}}. \quad (8)$$

This equation contains seven free parameters. For decision strategies that make different predictions for decision times and confidence for the considered item types (i.e., PCS, TTB) all seven parameters will be estimated. For strategies that predict equal decision times, the parameter R_T is not necessary (i.e., EQW, WADD) and hence only six parameters have to be estimated. Similarly, R_C can be omitted if a strategy makes all equal predictions for confidences. For a RAND strategy, R_T and R_C can be omitted as well as the error parameter ε_k which is set to be .50 (indicating random choices). Hence, for a RAND strategy only 4 parameters are estimated.

Likelihood values L_k should be corrected for the different numbers of free parameters N_p using the Bayesian Information Criterion (BIC) which also takes into account the number of observations N_{obs} (Schwarz, 1978):

$$BIC = -2\ln(L) + \ln(N_{obs})N_p. \quad (9)$$

Individuals should be classified as users of the strategy which has the lowest BIC value. The number of independent observations N_{obs}, which is used to calculate the BIC, is not always equal to the number of total observations. According to STATA 10.0 Online Manual, the number of independent categories (i.e., types of tasks) should be used if it can be assumed that the instances of these categories are highly correlated. This is the case for our data because responses to the repeated presentations of one type of items should be similar. I compared results using the total number of observations per person (N_{obs} = 60 [tasks]* 3[choice, decision time, confidence] = 180) and the number of independent categories (i.e., types) (N_{obs} = 6 * 3 = 18) in the simulation reported below and indeed found that the usage of the latter formula seems to be preferable.

The simulations investigated whether choices, decision times and confidence data generated by different strategies with certain error rates for choices and different effect sizes for decision time and confidence are correctly classified using the MM-ML method. I expected that this

---

[3]Alternatively, it could be assumed that σ differs between item types and increases with increasing t_i. Although this relation might also be modeled in ML calculation, for simplicity a constant σ should be assumed.

[4]The assumption that confidence judgments are normally distributed is rather common (e.g., Merkle, Sieck, & Van Zandt, 2008). For a discussion of the independence assumption, see below.

method a) is capable of identifying the decision strategy that generated the data, b) is not biased in favor of one or the other strategy, c) differentiates appropriately between strategies which make identical choice predictions (if the effect size for time and confidence is sufficiently large), and d) leads to less misclassifications than the usage of the choice-based strategy classification.

5 Simulation

5.1 Method

The simulation used the 5 decision strategies and 6 types of tasks from Table 1. I assumed that these tasks were presented 10 times each, resulting in a total of 60 choices. In the simulation, the choices, decision times and confidence were generated by the 5 strategies TTB, EQW, WADD, PCS, and RAND. The error rate for choices varied from 5% to 25% in steps of 5%. I also manipulated the size of the differences between decision times and confidences for different types of items in relation to the standard deviation. To do this, I drew data from normal distributions $N(\mu = contrast\ weight, \sigma = sd)$ in which the mean was the contrast weight defined in Table 1 and the standard deviation sd was varied on the levels 0.8, 1, 1.33, and 2. Remember that the contrast weights are scaled to a range of 1. Hence, $sd = 1$ means that for comparing the fastest with the slowest item type, the effect size is 1. The maximum effect sizes produced by our manipulation of sd are consequently 1.25, 1, 0.75 and 0.5. For simplicity, sd was manipulated jointly for decision times and confidences. For each combination from each strategy, 100 data sets were generated and the MM-ML strategy classification was applied. Hence, in the simulation I used a 5 (data-generating decision strategy) x 5 (error rate) x 4 (standard deviation) x 100 (repetitions) design. Simulations were run using a BIC correction with $N_{obs} = 18$ and $N_{obs} = 180$. The results for $N_{obs} = 18$ are reported only because they were consistently better (i.e., less biased in favor of strategies with less parameters).[5]

5.2 Results

Figure 1 shows the classification results by data generating strategy and maximal effect size (i.e., inverse of sd) aggregated over the manipulation of error rate. The classification for data generated by TTB and EQW were almost perfect. The classification of data generated by WADD was very good as well, although there was a small

[5]Following Schwarz (1978) we used BIC instead of the alternative Akaike information criterion AIC = -2*ln(Likelihood) + 2*N_p. Note, that using AIC in the simulations led to results similar to using $N_{obs} = 18$ except for the fact that strategies with more parameters were classified somewhat more often (because 2 < ln(18)).
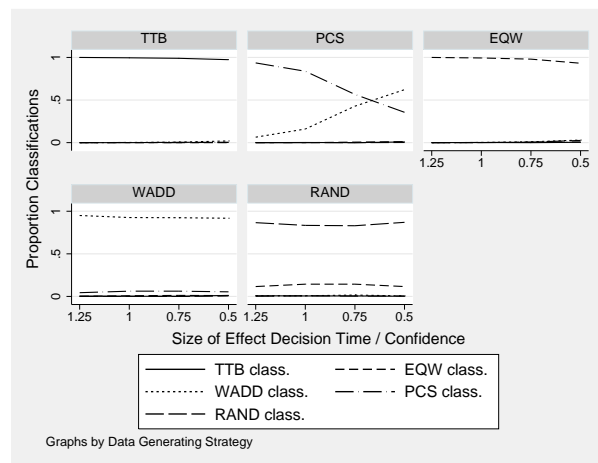

Figure 1: Strategy classification results by data generating strategy for 60 observations.

constant amount of misclassification in favor of PCS. Remember that WADD and PCS make equal choice predictions. Hence, the method generated very few misclassifications in favor of the more complex strategy (with one additional parameter). On the other hand, the accuracy of the classification of data produced by PCS depended crucially on the effect size. As one would expect, with decreasing effect size the number of misclassification in favor of the strategy not predicting a difference (i.e., WADD) increased. For our small number of observations, the maximal effect size (i.e., measured between the most extreme items only) of 1.25 and 1 led to acceptable results. Below that, misclassifications prevailed. Finally, data produced by a RAND strategy are to a certain degree misclassified as being produced by EQW. Note that this misclassification was likely to be due to the fact that, for the selected item types, EQW predicts random choice for 4 out of 6 considered types. These misclassifications could be reduced by including a limit error rate of .30 for all systematic strategies and not classifying participants with higher error rates (see discussion below).

The manipulation of the error rate in strategy application had only a minor influence on strategy classification results for all levels of sd. The results concerning the influence of error rate on strategy classification are shown in Figure 2. The left part shows the result aggregated for strong effects ($sd \leq 1$) and the right side for weaker effects ($sd > 1$). As one could have expected, for strong effects, strategy classification also worked quite well between PCS and WADD. The classification between these strategies with equal choice predictions was considerably worse if the effect was weaker. There was a strong tendency towards misclassification in favor of the strategy that does not predict differences in response times (i.e., WADD) as compared to the strategy that predicted differ-
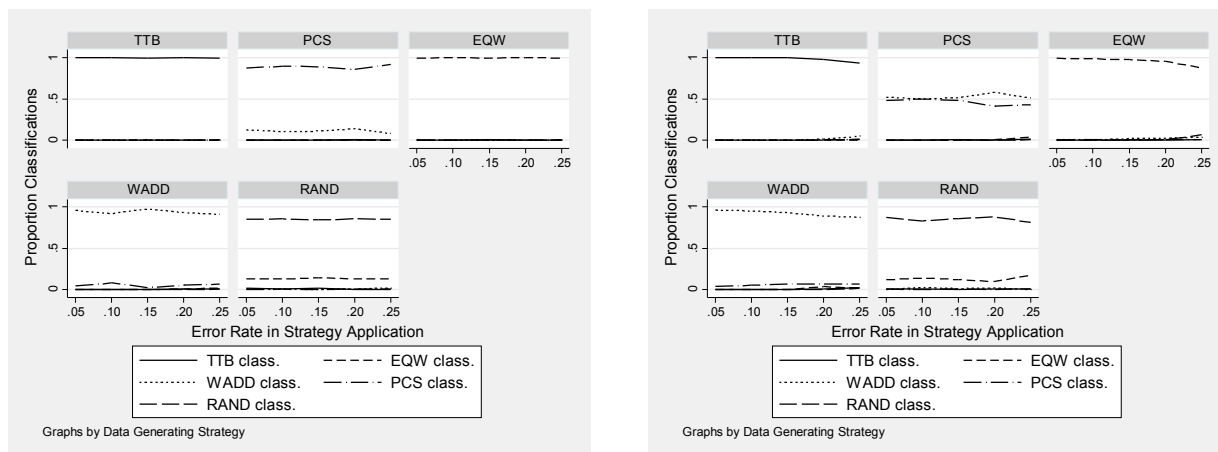
Figure 2: Strategy classification results by error rate in strategy application for strong effects (sd≤1, left) and weaker effects (sd>1, right).

ences (i.e., PCS). Hence with weaker effects the method is biased in favor of the strategy that predicts no difference (i.e., null hypothesis). In the high effect-size conditions, the increasing error rate had almost no increasing effect for misclassifications (Figure 2, left) whereas increasing error rate led to increasing misclassifications in the lower effect-size conditions (Figure 2, right).

To investigate whether the MM-ML method leads to fewer strategy misclassifications than the classic choice based strategy classification by Bröder and Schiffer (2003a), I rerun the same analysis using the choice based strategy classification method and excluding PCS (because it obviously could not be differentiated from WADD based on choices only). In line with findings by Bröder and Schiffer, the analysis worked very well, but revealed an increasing misclassification rate with an increasing error rate (Figure 3).

A considerable number of choices that were produced by RAND were wrongly classified as being produced by EQW. This bias was stronger as compared to the one observed for the MM-ML method (see Figure 1). In Table 2 the classification results for both methods are directly compared for $\varepsilon = 0.25$. It can be seen that over all strategies the MM-ML method leads to a higher level of correct classifications as compared to the choice based strategy classification (cf. bold numbers in the diagonals of Table 2).

A final simulation investigated the influence of the number of observations on the quality of the strategy classification with the MM-ML method. Therefore the number of observations used in the analysis was raised from 60 to 120 (i.e., by using 20 instead of 10 decisions per item type). Doubling the number of observation increased the quality of classification (Figure 4). With 120 observations the classifications of TTB, EQW, WADD,
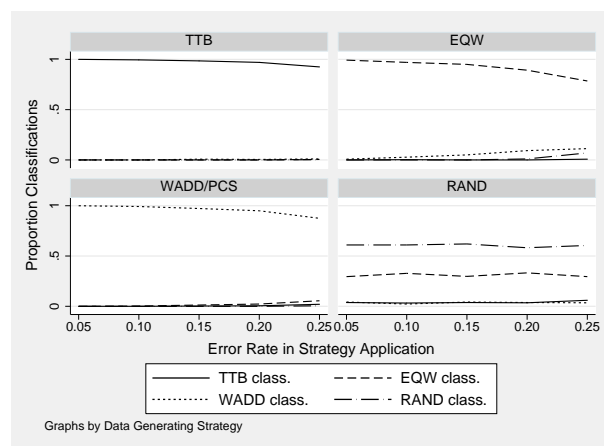

Figure 3: Strategy classification based on choices only by data generating strategy and error rate in strategy application.

and RAND were close to perfect. With the higher number of observations, the classification of PCS was also satisfactory for a lower maximum effect size (i.e., 0.75) but for the lowest maximum effect size (i.e., 0.5) there were still a considerable number of misclassifications in favor of WADD.

5.3 Discussion

The simulations revealed that the inclusion of decision times and confidences in the analysis generally improves strategy classification. This is particularly the case if the effects for both variables are strong. If the effects are strong, the method also allows differentiating reliably between strategies which make the same choice predictions. In cases with weaker effects, the method is increasingly

Table 2: Comparison of strategy classification methods.

Data generating strategy	Strategy classification			
	TTB	EQW	WADD/PCS	RAND
Choice-based strategy classification				
TTB	**0.93**	0.01	0.01	0.00
EQW	0.01	**0.79**	0.11	0.07
WADD/PCS	0.02	0.06	**0.88**	0.01
RAND	0.06	0.30	0.04	**0.61**
Overall	1.01	1.14	1.03	0.68
MM-ML strategy classification				
TTB	**0.96**	0.00	0.03	0.01
EQW	0.01	**0.94**	0.03	0.03
WADD/PCS	0.01	0.02	**0.96**	0.01
RAND	0.00	0.15	0.02	**0.83**
Overall	0.98	1.10	1.04	0.88

Note. Numbers represent percentages of strategy classification for the respective strategy (in columns) for an error rate of $\varepsilon = .25$ only. In the MM-ML method, WADD and PCS are combined concerning data generating strategy and strategy classification for comparative reasons.

biased in favor of the strategies which predict no difference concerning decision times and choices (i.e., which has less free parameters). This problem obviously occurs with any statistical test because the latter strategies have the advantage of being the null hypothesis.

The simplest way to circumvent the problem is to include items for which particularly strong differences in confidence and decision time are expected. Additionally, more items could be used to increase power by increasing within-subjects sample-size. An increase from 10 to 20 choices per item type reduces the bias in favor of the null hypothesis considerably. Finally, one might consider using a different correction of the likelihood than the BIC correction (similar to setting a compromise alpha level). However, to the best of my knowledge there is no simple method to find the correct adjustment (although it could, of course, be derived from simulations). Hence including items with expected large differences and increasing the number of items seem to be preferable. Note that previous studies found strong effects for confidence and time in probabilistic inference tasks (Glöckner, 2006; Glöckner & Betsch, 2008c), as well as in gambling tasks (Glöckner & Betsch, 2008a; Glöckner & Herbold, 2008).
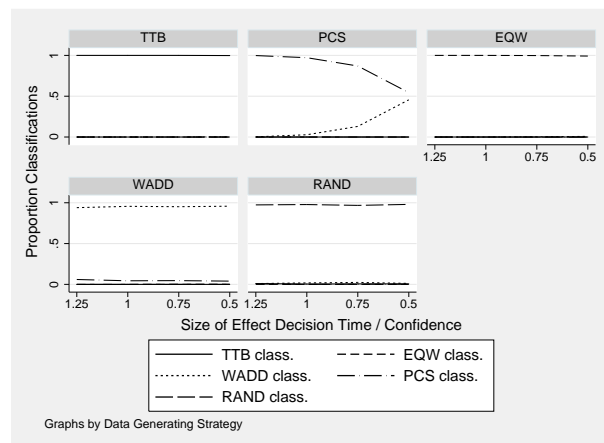

Figure 4: Strategy classification results by data generating strategy for 120 observations.

These findings indicate that biased classification due to weak effect size might not be too much of a problem. However, researchers should check the size of the effect in pre-tests, or they should at least calculate it before interpreting their results.

6 Applying the MM-ML method in research practice

Applying the MM-ML method obviously necessitates the use of a statistical package that allows for calculating complex maximum likelihood estimations. I have programmed the necessary estimation routines in STATA. The estimation programs are described in the supplementary material (http://journal.sjdm.org/vol4.3.html). Applying the method mainly requires bringing data into a specific format and defining predictions. The overall estimation program (which can be applied to any number of item types, choices per item, participants, and strategies) provides per-individual estimates of the parameters for each strategy (Figure 5, top), as well as an aggregated matrix with the total likelihood (i.e., BIC score) that the data for each individual (in rows) were produced by a particular strategy (in columns) (Figure 5, bottom).

The STATA output will be explained in more detail. The individual output (Figure 5, top) shows the results for comparing the data of subject 1 with the predictions of strategy 4 (see last line of output). The total number of observations is 126 (6 frequencies for choices in task types, 60 decision times, 60 confidence ratings). The resulting parameter estimates are listed as constant coefficients. In the example, the choice error rate (*epsilon*) was .167, the log-transformed (and for order ef-

```
Iteration 8:    log likelihood = -332.38404
Iteration 9:    log likelihood = -332.38404

                                            Number of obs  =        126
                                            Wald chi2(0)   =          .
Log likelihood = -332.38404                 Prob > chi2    =          .
```

dv	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
epsilon					
_cons	.1666667	.0481124	3.46	0.001	.0723681 .2609654
mu_Time					
_cons	8.572725	.0398664	215.04	0.000	8.494588 8.650862
sigma_Time					
_cons	.3088002	.0281895	10.95	0.000	.2535499 .3640506
R_Time					
_cons	.3235926	.1202474	2.69	0.007	.087912 .5592731
mu_Conf					
_cons	53.98333	4.288501	12.59	0.000	45.57803 62.38864
sigma_Conf					
_cons	33.21859	3.032428	10.95	0.000	27.27514 39.16204
R_Conf					
_cons	70.48136	12.16382	5.79	0.000	46.64072 94.32201

```
Subj: 1  Strat: 4    - BIC:685.00069


. matrix list result

result[10,5]
            c1          c2          c3          c4          c5
 r1   678.6011   741.32443   685.13987   685.00069   738.94167
 r2   635.65525   608.39091   592.08052   594.65917   652.84004
 r3   684.55589   710.90953   642.06809   636.01209   723.15714
 r4   708.55863   735.07529    643.3949   634.74853   756.18883
 r5    694.497    689.10911    613.9352   615.16598   716.39535
 r6   675.90311   716.37188   655.40301   649.69735   734.16568
 r7    696.3077   696.26612   645.39747   643.47576    749.5148
 r8    702.0289   766.19389   711.55809   709.12233   763.81113
 r9   648.61714   642.35812   594.47847   592.13334   669.56801
r10   570.06646   669.29671   600.18226   601.89131   666.93234
```

Figure 5: Example output of the STATA implementation of the Multiple-Measure Maximum Likelihood strategy classification method for parameters per individual (top) and for the overall estimation (bottom). The individual output contains estimates for all coefficients and the overall fit of the individual data to the prediction of the considered strategy. The overall estimation shows BIC scores for each individual (rows) and each of the five considered strategies (columns). Lower scores indicate a better fit.

fects corrected) mean decision time was 8.57 (*mu_Time*), the mean confidence (*mu_Conf*) was 53.98. The provided significance tests (which test if the estimated constant coefficient is different from zero) are mainly informative for the rescaling factors *R*. In this example, R_T (*R_Time*) and R_C (*R_Conf*) were both significantly different from zero. This indicates that the specific predictions for time and confidence (reflected in contrast weights) significantly contribute to explain the data. The tests for *R* produce results similar to correlations between data and contrast weights (i.e., the correlations of observed decision times with the contrast weights for the respective task types).[6] The BIC score (last line) indicates the overall fit of data and strategy predictions for the specific participant. More precisely, it gives the corrected log-transformed likelihood for the data given the application of a certain strategy under the assumptions of a constant error rate ε, nor-

mally distributed decision times and confidence ratings, and independence between observations.

The lower part of Figure 5 shows the output for the results of all individual comparisons. It presents the resulting BIC scores for each subject (in rows) and considered strategy (in columns; i.e., TTB, EQW, WADD, PCS, RAND). Lower values indicate a better fit. The example result of subject 1 and strategy 4 is consequently presented in row 1, column 4 of the matrix. The strategy that explains a subjects' data best can be easily determined by identifying the lowest number in the persons' row (e.g., the data of subject 1 are most likely generated by strategy 1).

The MM-ML method has been successfully applied to empirical data (i.e., Figure 5 is based on real data) and it appears that for the types of items considered here (using 60 observations only) the method is well applicable. Additional practical suggestions on the application of the method are given in Glöckner (in press).

---

[6]Differences result from the fact that the parameter is estimated jointly with the other parameters in the MM-ML method.

7 Limitations of the method and suggested solutions

7.1 Exhaustive Set of Strategies

The quality of strategy classification is crucially dependent on the set of competing strategies that are considered. As pointed out by Bröder and Schiffer (2003a), one has to assume that (optimally) an exhaustive set of strategies is investigated. Although the selection of an exhaustive set of strategies is practically not possible, the suggested statistical method (and also the provided STATA program) can be used to compare any number of strategies. Pragmatically, researchers should nevertheless aim to consider only plausible strategies to hold the analysis manageable.

7.2 Criterion for classification of strategy use

Absolute value. One of the frequently raised questions for ML-strategy classification is whether there should be a criterion of fit that has to be reached in order to classify a person as having used a systematic strategy. Could an absolute likelihood or BIC value (e.g., BIC < 500) be defined that has to be reached for a classification? Considering equation 8 the answer has to be no. The total likelihood and also the BIC score crucially depend on the number of observations considered. The BIC score increases (and the likelihood decreases) with the amount of considered observations because more likelihood values (which are usually smaller than 1) are multiplied with each other.

Error in choices. The simplest (advisable) criterion for non-classification is to use a criterion that determines the maximal acceptable error rate in choices for systematic strategies. The lowest useful criterion is .50 (chance level). Considering the simulation results, a somewhat stricter criterion of $\varepsilon < .30$ could be advisable for low numbers of observations. If a researcher has good reason to believe that participants make only few errors (i.e., incentivized environment with clearly structured information), even more strict criteria might be used. The aspired limit error rate can be easily changed in the STATA estimation program (see supplementary material). Setting a stricter error rate limit for all systematic strategies increases the number of cases in which persons are classified as users of RAND. Note that a RAND strategy should always be included in the strategy classification. This assures that for strategy classification a systematic strategy has to be better than a random choice after correcting for the additional free parameters.

Bayes ratio. Another possibility is to compare likelihoods by determining the Bayes ratio. The Bayes ratio is calculated by dividing the likelihood value of the most likely strategy by the likelihood of the second most likely strategy (Wasserman, 1999). The reliability of a classification increases with increasing Bayes ratio. According to Wasserman (1999) ratios larger than 3 can be considered moderate evidence for the model which might be considered a lower limit for strategy classification. Note however that the Bayes ratio is calculated on likelihoods that are not corrected for number of free parameters. Hence, comparing a strategy which makes differential predictions for choices, times and confidence (e.g., PCS) with a strategy which does not (e.g., RAND) would lead to biased Bayes ratios. The application of Bayes ratios as criterion for classification vs. non-classification is therefore in a MM-ML method often not possible.

7.3 Independence of observations assumption

As mentioned above, the multiplication of likelihoods used in the MM-ML method (i.e., equation 8) relies on the assumption that likelihood values for choice, decision times and confidence are independent of each other. Considering the finding that decision time and confidence are often negatively correlated (e.g., Glöckner & Betsch, 2008c), this assumption might seem questionable at first glance. A closer look, however, reveals that such a correlation does not challenge the independence assumption underlying the MM-ML method because the method takes into account correlations that are predicted by strategies. In Table 1 it can be seen that the correlation between confidence and time should be $r = -1$ according to TTB and $r = -.85$ according to PCS (assuming that all 6 cue combinations are equally likely). Remember that likelihoods are calculated based on the assumption that values (for time and confidence) are normally distributed around the predicted mean for the respective item type (see equation 6). Therefore likelihoods are based on deviations from the mean *after* correcting for systematic differences in means (i.e., correlations of residuals after partialling out systematic effect of item types).

Furthermore, note that in the simulations reported above these systematic correlations between time and confidence were also induced to the data by generating them from the predictions of the strategies (Table 1). The size of the correlation was implicitly manipulated by adding relatively small or large error terms to these systematic components (i.e., increasing correlation with increasing maximal effect size). Hence, the simulations also show that these correlations do not lead to biases in strategy classification.

7.4   Correction for learning effects

It can be expected that decision times decrease over time, particularly if persons have to repeat each item type 10 or more times. This could harm strategy classification based on decision time. To avoid systematic biases induced by order, an individually randomized presentation order should be used. Furthermore, it is advisable to reduce error variance by partialling out the effect of order on log-transformed decision times and to use the resulting residuals instead of the raw values in the MM-ML method.

7.5   Dependence of parameters and estimation of strategies with additional parameters

Another possible caveat for the method might be that the different estimated parameters are not mutually independent. For instance, it might be argued that σ increases with increasing μ (i.e., heteroscedasticity). This can, however, be easily handled in STATA by including one parameter as predictor for the other in the estimation program (Gould, Pitblado, & Sribney, 2006). Before doing so, it is necessary, however, to have a good hypothesis about the relation of the parameters. Furthermore, one might want to test strategies that have free parameters themselves (e.g., Bergert & Nosofsky, 2007; Busemeyer & Johnson, 2004; Busemeyer & Townsend, 1993; Nosofsky & Bergert, 2007). This can, of course, also be handled by including these strategy-parameters in the estimation. Finally, if strategies make predictions on only one of the continuous variables the method can be applied in a simplified version (as indicated in equation 7).

8   General discussion

The Multiple-Measure Maximum Likelihood strategy classification method allows for identifying individuals' decision strategies by taking into account choices, decision times and confidence judgments at the same time. In contrast to earlier approaches for including decision time into strategy classification (Bergert & Nosofsky, 2007; Glöckner, 2006; for an overview, see Glöckner, in press), it allows estimating the overall likelihood of the data given the application of a strategy and comparing any number of strategies based on the whole set of observations. The method allows differentiating between strategies that make the same choice predictions as long as their effects on decision time and confidence are sufficiently large. With decreasing effect sizes and small numbers of observations, the method is biased towards the strategy with fewer parameters. It is therefore advisable to

use types of items for which large differences in times and confidences are expected, to use sufficient items per item type and to check the effect size before interpreting the results concerning strategies which make equal choice predictions.

It could be shown that including decision times and confidences reduces the proportion of misclassification as compared to the choice-based strategy classification by Bröder and Schiffer (2003a). Therefore, it is advisable to use the MM-ML method even in cases where different choice predictions can be derived from different strategies.

Besides providing an overall maximum likelihood measure, the MM-ML method provides a tool to improve our understanding of the processes underlying decision making. It allows investigating the fit for the different dependent variables separately for each strategy and each individual. As can be seen in Figure 5 (left), the method provides a significance test for the scaling parameters R_C and R_T, which indicates whether the specific prediction for confidence or decision time of the respective strategy was in line with the data or not. Process models might be improved based on this knowledge by simply counting the number of significant predictions for each dependent variable.

8.1   Further applications and extensions

The MM-ML method is very general and not limited to investigate probabilistic inference tasks or to decision research in general. It can be applied for testing any kind of models for cognitive processes that make predictions concerning dichotomous behavior, response times and confidences (e.g., gambling tasks, recognition tasks, multi-attributive decisions). For example, recent research on strategy selection in gambling decisions (e.g., Glöckner & Betsch, 2008a) could be extended by recording choices, decision times and confidence and comparing the total data with predictions of different models (e.g., prospect theory, decision field theory, PCS, priority heuristic). The method can, of course, also be extended to further (normally distributed) dependent measures which can basically be included in the analysis in the same way as decision time and confidence (see above). A possible extension of particular interest to those investigating intuition could be measures for the distribution of eye-fixations and physiological arousal which can both be captured by recent eye-tracking technology (Glöckner & Herbold, 2008; see also Hochman, Glöckner, & Yechiam, in press). For the investigation of deliberate strategies only, classic mouselab measures for information search such as the Payne index (Payne et al., 1988) could also be potentially included in the analysis.

In sum, the MM-ML method is an easy to use and reliable method for individual level strategy classification, which can be applied for intuitive and deliberate strategies. Application of the method, however, necessitates well specified models (Glöckner, in press). The MM-ML method has the major advantage to test strategies or models considering all their predictions at the same time. Furthermore, it will help to improve process models by providing much more detailed information on the fit of predictions for different dependent measures.

References

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

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

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica, 87*, 137–154.

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*, 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. (in press). Outcome-based strategy classification. In A. Glöckner & C. L. M. Witteman (Eds.), *Tracing intuition: Recent methods in measuring intuitive and deliberate processes in decision making.* London: Psychology Press & Routledge.

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

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

Busemeyer, J. R., & Johnson, J. G. (2004). Computational models of decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 133–154). Malden, MA: Blackwell Publishing.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432–459.

Doherty, M. E., & Brehmer, B. (1997). The paramorphic representation of clinical judgment: A thirty-year retrospective. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections, and controversies* (pp. 537–551). New York: Cambridge University Press.

Doherty, M. E., & Kurz, E. M. (1996). Social judgment theory. *Thinking & Reasoning, 2*, 109–140.

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.

Fishburn, P. C. (1974). Lexicographic orders, utilities, and decision rules: A survey. *Management Science, 20*, 1442–1472.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

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

Glöckner, A. (2006). *Automatische Prozesse bei Entscheidungen [Automatic processes in decision making]*. Hamburg, Germany: Kovac.

Glöckner, A. (2007). Does intuition beat fast and frugal heuristics? A systematic empirical analysis. In H. Plessner, C. Betsch & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 309–325). Mahwah, NJ: Lawrence Erlbaum.

Glöckner, A. (in press). Multiple measure strategy classification: Outcomes, decision times and confidence. In A. Glöckner & C. L. M. Witteman (Eds.), *Tracing Intuition: Recent Methods in Measuring Intuitive and Deliberate Processes in Decision Making*. London: Psychology Press & Routledge.

Glöckner, A., & Betsch, T. (2008a). Do people make decisions under risk based on ignorance? An empirical test of the Priority Heuristic against Cumulative Prospect Theory. *Organizational Behavior and Human Decision Processes, 107*, 75–95.

Glöckner, A., & Betsch, T. (2008b). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making, 3*, 215–228.

Glöckner, A., & Betsch, T. (2008c). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1055–1075.

Glöckner, A., & Herbold, A.-K. (2008). Information processing in decisions under risk: Evidence for compensatory strategies based on automatic processes. *MPI Collective Goods Preprint, No. 42. Available at SSRN:* `http://ssrn.com/abstract=1307664`.

Glöckner, A., & Witteman, C. L. M. (in press). Foundations for tracing intuition: Models, findings, categorizations. In A. Glöckner & C. L. M. Witteman (Eds.), *Tracing intuition: Recent methods in measuring intuitive and deliberate processes in decision making*. London: Psychology Press & Routledge.

Gould, W., Pitblado, J., & Sribney, W. (2006). *Maximum Likelihood Estimation with Stata* (3rd ed.). College Station, TX: Stata Press.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, & Cybernetics, 17*, 753–770.

Harte, J. M., & Koele, P. (2001). Modelling and describing human judgement processes: The multiattribute evaluation case. *Thinking and Reasoning, 7*, 29–49.

Hochman, G., Glöckner, A., & Yechiam, E. (in press). Physiological measures in identifying decision strategies. In A. Glöckner & C. L. M. Witteman (Eds.), *Tracing intuition: Recent methods in measuring intuitive and deliberate processes in decision making*. London: Psychology Press / Routledge.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57*, 116–131.

Horstmann, N., Ahlgrimm, A., & Glöckner, A. (under review). How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes.

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

Merkle, E. C., Sieck, W. R., & Van Zandt, T. (2008). Response error and processing biases in confidence judgment. *Journal of Behavioral Decision Making, 21*, 428–448.

Norman, E., & Schulte-Mecklenbeck, M. (in press). Take a quick click at that! Mouselab and Eye-Tracking as tools to measure intuition. In A. Glöckner & C. L. M. Witteman (Eds.), *Tracing intuition: Recent methods in measuring intuitive and deliberate processes in decision making*. London: Psychology Press / Routledge.

Nosofsky, R. M., & Bergert, F. B. (2007). Limitations of exemplar models of multi-attribute probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 999–1019.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534–552.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In *Simple heuristics that make us smart* (pp. 141–167). New York, NY: Oxford University Press.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

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