Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2015)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1878

Verbalization of Decision Strategies in Multiple-Cue Probabilistic Inference

MATTHEW M. WALSH* and KEVIN A. GLUCK
Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

ABSTRACT

In multiple-cue probabilistic inference, people choose between alternatives based on several cues, each of which is differentially associated with an alternative's overall value. Various strategies have been proposed for probabilistic inference (e.g., weighted additive, tally, and take-the-best). These strategies differ in how many cue values they require to enact and in how they weight each cue. Do decision makers actually use any of these strategies? Ways to investigate this question include analyzing people's choices and the cues that they reveal. However, different strategies often predict the same decisions, and search behavior says nothing about whether or how people use the information that they acquire. In this research, we attempt to elucidate which strategies participants use in a multiple-cue probabilistic inference task by examining verbal protocols, a high-density source of process data. The promise of verbal data is in their utility for testing detailed information processing models. To that end, we apply protocol analysis in conjunction with computational simulations. We find converging evidence across outcome measures, search measures, and verbal reports that most participants use simplifying heuristics, namely take-the-best. Copyright © 2015 John Wiley & Sons, Ltd.

Additional supporting information may be found in the online version of this article at the publisher's web-site.

KEY WORDS multiple-cue probabilistic inference; verbal protocols; process measures; take-the-best; tally; weighted additive

Traditional utility theories postulate a decision maker with unlimited time and mental resources available to perform calculations. The decision maker also has complete knowledge of available actions, as well as the likelihoods of their associated outcomes. In other words, the decision maker is omnipotent and omniscient. Within this framework, rationality is defined by the correspondence between the decision maker's choices and the laws of logic and probability (Edwards, 1954; von Neumann & Morgenstern, 1944). This view ignores inherent limitations that real decision makers face. In the real world, choices must be made quickly, using finite mental resources, and with incomplete information.

An alternate view is that people use simple heuristics rather than formal analyses to make decisions (Gigerenzer, Todd, & ABC Research Group, 1999; Payne, Bettman, & Johnson, 1988; Simon, 1955). Heuristics are fast because they rely on simple mental operations, and they are frugal because they require little information to enact (Gigerenzer & Gaissmaier, 2011; Shah & Oppenheimer, 2008). Additionally, by capitalizing on statistical regularities in the environment, heuristics can perform as well as—or better than—methods that require complex calculations and greater amounts of information (Gigerenzer & Brighton, 2009). Within this framework, rationality is defined by the correspondence between the decision maker's choices and the structure of the environment (Gigerenzer & Goldstein, 1996; Hogarth & Karelaia, 2007; Simon, 1955). When the decision strategy and the environment are well aligned, the choice is ecologically rational.

Gigerenzer and colleagues proposed that the mind contains an "adaptive toolbox" equipped for solving a multitude of problems (Gigerenzer et al., 1999; Gigerenzer & Gaissmaier, 2011). This toolbox is composed of heuristics, their building blocks (e.g., search rules, stopping rules, and decision rules), and the cognitive capacities that they depend on (e.g., associative memory). Although this view has been tremendously influential, it is also controversial. One key debate concerns the strength of empirical evidence for the use of heuristics (Chater, Oaksford, Nakisa, & Redington, 2003; Dougherty, Franco-Watkins, & Thomas, 2008; Glöckner & Betsch, 2010; Hilbig & Richter, 2011; Newell, 2005). The question of whether heuristics are fast, frugal, and accurate is separate from the question of whether they are adequate models of the decision strategies that people actually use. In this article, we examine the latter question. We focus on one type of problem, probabilistic inference, and on one of the heuristics thought to be in the adaptive toolbox, take-the-best (TTB). Given the centrality of TTB to the adaptive toolbox theory (Gigerenzer & Goldstein, 1996), it is important to ask, do people actually use TTB and how can we know?

Multiple-cue probabilistic inference

In multiple-cue probabilistic inference, people decide which of the two alternatives has greater value for a specified criterion based on several cues, each of which is differentially associated with an alternative's overall value.[1] For example, an investor might consider multiple financial indicators before deciding which of the two stocks to buy. These types of decisions are complicated by two factors. First, no single cue or combination of cues typically predicts the correct alternative perfectly; outcomes are probabilistic. Second, different cues

*Correspondence to: Matthew M. Walsh, Air Force Research Laboratory, 711 HPW/RHAC–Cognitive Models and Agents Branch, 2620 Q Street, Building 852, Wright-Patterson Air Force Base, OH 45433, USA. E-mail: mmw188@gmail.com

[1]In the related problem of multi-attribute choice, people also choose between alternatives based on multiple attributes. The distinction between probabilistic inference and multi-attribute choice is that the decision criterion is objective in the former but subjective in the latter. Still, inferences and preferences may share some cognitive processes (Weber & Johnson, 2009).

typically favor different alternatives; no alternative is dominant.

Several strategies have been proposed for performing multiple-cue probabilistic inference. Weighted additive (WADD) computes the sum of cue values multiplied by the weight of each cue and selects the alternative with the greatest overall value (Payne et al., 1988). WADD is not fast; it requires complex mathematical calculations. Additionally, WADD is not frugal; it requires accessing all cue values every time a decision is made. Still, WADD is a statistically motivated, normative solution to multiple-cue probabilistic inference.

Other simpler approaches exist. Tally (TAL) ignores weights and simply counts the number of positive cues for each alternative (Dawes, 1979; Einhorn & Hogarth, 1975). TAL is fast because it uses simple mathematical operations (i.e., counting). However, TAL still requires accessing all cue values every time a decision is made. A final approach, TTB, searches cues in order of their validity, stops upon identifying a cue that distinguishes between alternatives, and selects the alternative with the greater cue value (Gigerenzer & Goldstein, 1996).[2] TTB is fast because it uses simple mathematical operations (i.e., binary comparison), and it is frugal because it can be applied given very little information about alternatives. Surprisingly, simulation studies show that despite its simplicity, TTB can perform as well as—or better than—WADD (Czerlinski, Gigerenzer, & Goldstein, 1999; Gigerenzer & Brighton, 2009; Martignon & Hoffrage, 1999). Specifically, TTB is most accurate when a single cue is more important than any combination of less valid cues (i.e., a noncompensatory environment as opposed to a compensatory environment; Martignon & Hoffrage, 1999).

Empirical tests of take-the-best

Although statistical analyses and computer simulations have demonstrated the ecological rationality of TTB, they do not establish that people actually use this heuristic. As asserted by Newell and Shanks (2003), "… such simulation data indicate nothing about TTB's adequacy as a description of actual human behavior" (p. 54). Empirical results that bear on this issue are mixed, as we describe next.

Outcome measures

One way to distinguish among strategies is to examine the outcome of the decision process. Under certain conditions, the majority of decisions are consistent with the use of TTB. For example, TTB better predicts people's choices in noncompensatory environments than in compensatory environments (Bröder, 2000, 2003; Rieskamp & Otto, 2006). Additionally, participants favor TTB over WADD when information acquisition is costly (Bröder, 2000; Newell & Shanks, 2003; Rakow, Newell, Fayers, & Hersby, 2005) and when time pressure is great (Glöckner & Betsch, 2008; Rieskamp & Hoffrage, 2008). Presumably, this is because of TTB's less exhaustive search, which makes it faster and more frugal.

Other outcome-based results are less consistent with TTB. Several studies have found that choices are influenced by the values of cues with low validity (Ayal & Hochman, 2009; Bröder, 2000; Söllner, Bröder, Glöckner, & Betsch, 2014). This would not be expected if people only considered the most important discriminating cue. Additionally, even when conditions decidedly favor TTB, the selections of a substantial minority of participants are not consistent with TTB (Bröder, 2000; Glöckner & Betsch, 2008; Newell & Shanks, 2003).

Process measures

Another way to distinguish among decision strategies is to focus on the decision process itself rather than its outcome (for a review, see Schulte-Mecklenbeck, Kühberger, & Ranyard, 2011). Mata, Schooler, and Rieskamp (2007) asked participants to choose among alternatives displayed in a table on a computer screen. To view cue values, participants selected the corresponding cells in the table using a cursor. They acquired less information when cues were noncompensatory (for a related example in the domain of risky choice, see Payne et al., 1988). People also view fewer cues when information acquisition is costly and when time pressure is great, the very conditions that favor TTB (Bröder, 2000; Rakow et al., 2005; Rieskamp & Hoffrage, 2008). Lastly, when making inferences from memory, participants' reaction times increase monotonically with the number of cues that must be retrieved before finding one that discriminates between alternatives, consistent with the use of TTB (Bröder & Gaissmaier, 2007).

Not all process-based results support TTB, however. People typically acquire more information than is needed to use TTB even when their choices are ultimately consistent with that strategy (Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). Reisen, Hoffrage, and Mast (2008) asked participants to retrospectively describe their decision strategies. Although participants' choices were consistent with their reports, they acquired about 50% more information than their strategies called for. Finally, the elapsed time before a decision is made is sometimes better predicted by the degree of coherence among cues rather than the minimum number that must be examined to use TTB (Glöckner & Betsch, 2012).

Limitations

Although informative, the outcome and process measures typically used to study probabilistic inference do have limitations. First, the choices predicted by TTB, WADD, and TAL overlap substantially (Bröder, 2003), rendering decisions inconclusive as evidence for process. This problem is compounded by the fact that participants may apply rules inconsistently between trials. Second, the process measures predicted by TTB also overlap with those predicted by other decision strategies (Harte, Westenberg, & van Someren, 1994; Reisen et al., 2008). Third and finally, search behavior

[2]Payne et al. (1988) described the related lexicographic heuristic for preference judgments. Although they never spoke of TTB, it is a special case of the lexicographic heuristic.

DOI: 10.1002/bdm

shows what information people acquire and the order in which they do so but not whether or how they use that information. This has led some to question the correspondence between information search and integration (Bröder, 2003; Harte & Koele, 2001; Maule, 1994; Rieskamp & Hoffrage, 2008). In sum, outcome measures neglect important details about predecisional behavior, and process measures provide limited insight into how people integrate information. What is needed is a method that provides detail about whether and how people use the information that they acquire to make decisions.

Verbal protocols

To overcome these limitations, we used verbal protocol analysis to study the cognitive processes underlying multiple-cue probabilistic inference. At the heart of protocol analysis is the idea that people can verbalize thoughts and that controlled, purposive mental operations can be inferred from verbalizations (Ericsson & Simon, 1993). The heuristics that people are thought to use in probabilistic inference have been formalized as computational models (Gigerenzer & Gaissmaier, 2011). These process models predict the content of participants' verbalizations and their overt behaviors. As such, verbal protocol analysis is an appropriate and overdue methodology to incorporate into the scientific study of multiple-cue probabilistic inference.

Although verbal protocols have not yet been used to study heuristics in multiple-cue probabilistic inference, they have been applied to other topics in judgment and decision making research (Brandstätter & Gussmack, 2013; Cokely & Kelley, 2009; Tenbrink & Wiener, 2009). For example, Brandstätter and Gussmack (2013) recorded verbal reports from participants as they chose between gambles. They found that participants predominantly compared individual attributes between gambles (e.g., "Gamble A has a *higher probability* of winning than Gamble B" or "Gamble A has a *lower gain* than Gamble B") rather than forming separate, overall impressions of each gamble first (see also Cokely & Kelley, 2009; Johnson & Schkade, 1989). Verbal protocols have also been used to examine preferences in multi-attribute consumer choice (for reviews, see Harte et al., 1994; Ranyard & Svenson, 2011). Based on verbal data, Payne (1976) concluded that as problem complexity increased, participants began to favor noncompensatory decision strategies that quickly eliminated alternatives from the consideration set (see also Johnson & Meyer, 1984; Timmermans, 1993). These examples demonstrate the value of using verbal data to study heuristics in judgment and decision making.

Overview of present experiment

To address the question of whether people use TTB, we recorded verbal reports from participants as they solved multiple-cue probabilistic inference problems. If they used a noncompensatory strategy like TTB, we expected that their statements would predominantly involve comparing individual cues between alternatives and selecting an alternative based on one cue. Alternatively, if they used a compensatory strategy like WADD or TAL, their statements would involve combining and comparing information from multiple cues.

We also tested the ancillary hypothesis that verbalizing would not affect the cognitive processes involved in multiple-cue probabilistic inference. Verbal reporting is not generally thought to cause reactivity—differences between how people perform a task silently and while thinking aloud (Fox, Ericsson, & Best, 2011). Yet this method remains controversial because of the concern that it may affect the cognitive processes being investigated (Schooler, 2011). Thus, the second aim of this study was to determine whether verbalization caused reactivity in multiple-cue probabilistic inference.

METHOD

Participants

Thirty-eight individuals from the University of Dayton each participated in a 1-h session for monetary compensation (19 women, ages ranging from 18 to 34 years with a mean age (±1 standard deviation) of 22 ± 4 years). Participants were randomly assigned to one of the two experiment conditions: a verbalization condition in which they thought aloud while performing the task and a silent condition in which they did not.

Task

Participants completed a simulated stock-market selection task that has been used before to study multiple-cue probabilistic inference (Bröder, 2000; Newell et al., 2003). In each trial, they chose between two hypothetical stocks (Figure 1). Participants could view four indicators about each stock: (i) established company, (ii) invest in new projects, (iii) financial reserves, and (iv) share trend positive. To reveal the value of an indicator for a stock, they used a computer cursor to select the corresponding cell in the experiment interface. After they clicked a cell, the value "yes" appeared (i.e., yes, company IFH has financial reserves) or the value "no" appeared (i.e., no, company EUI does not have financial reserves). The interface became temporarily inactive for 1 s after an indicator was revealed. This allowed us to collate verbal reports with specific actions. To choose a stock, participants clicked the corresponding button below the grid. The name of the correct stock appeared in the box labeled as best stock, trial pay was displayed, and total pay was updated.

Trial pay depended on two factors. First, participants received 10 ¢ for selecting the winning stock and 0 ¢ otherwise. Second, participants paid 1 ¢ to view each indicator. To maximize trial pay, they needed to view enough indicators to make informed decisions without spending too much of acquiring information. The payoff structure of the task was explained to participants before the experiment began.

Cue validities were displayed beside the four indicators and were freely visible throughout the experiment. The order of indicators on the screen was randomized across participants, as was the assignment of validities to each indicator (.80, .75, .70, and .69). Cue validities were identical to values

DOI: 10.1002/bdm

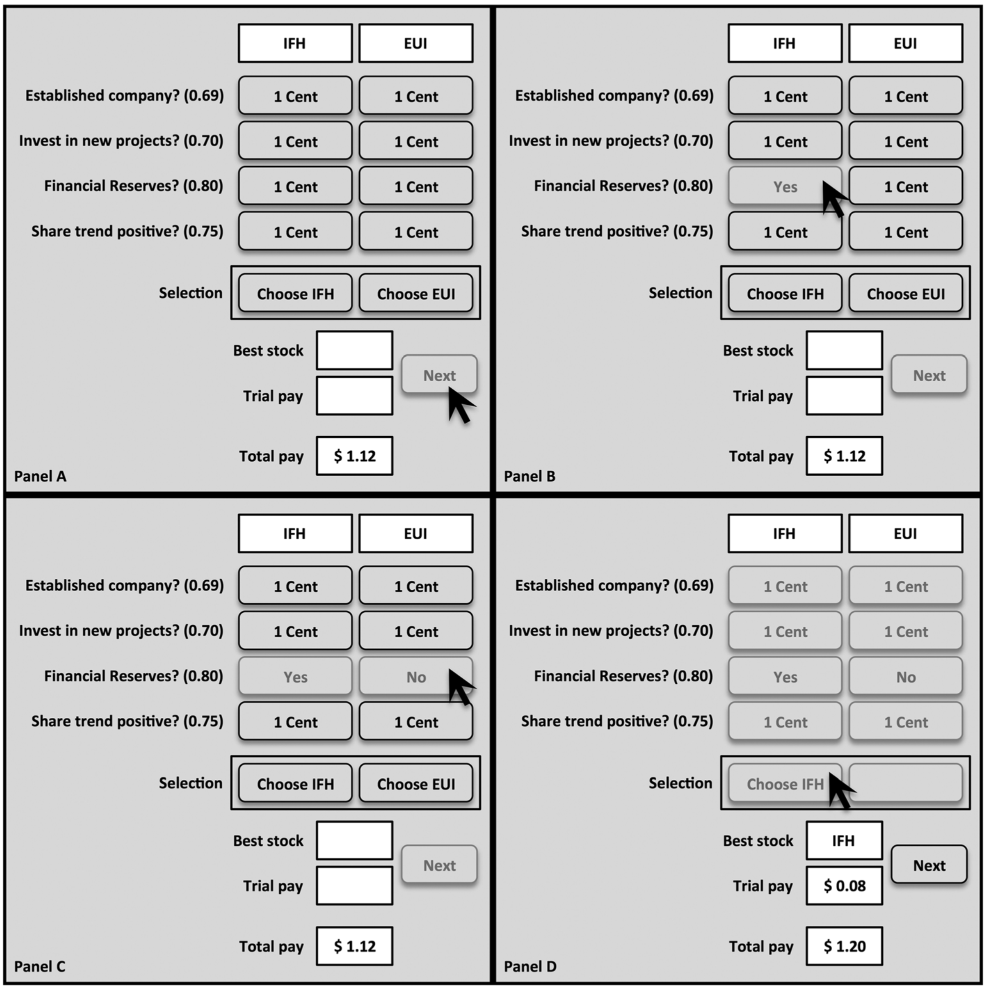

Figure 1. Experiment interface. Stock names appeared at the top of the screen, and indicator labels and validities were displayed along the left-hand side of the screen. Indicator values were concealed in the grid below the stock names (Panel A). After the participant clicked an indicator for a stock, the value "yes" or "no" appeared in the corresponding cell (Panels B and C). To choose a stock, the participant clicked the stock's button below the grid. The name of the best stock and trial pay then appeared (Panel D)

that have been used before in multiple-cue probabilistic inference experiments (Bröder, 2000; Newell et al., 2003). The meaning of cue validities was explained to participants before the experiment began.

The experiment contained a total of 120 trials. Cue configurations were created such that TTB and WADD predicted different choices from one another in 20 trials (16%) and that WADD and TAL predicted different choices from one another in 30 trials (25%).

Procedure

After providing informed consent, participants were given instructions about verbal protocols (Ericsson & Simon, 1993). They were told,

> I will ask you to think aloud as you work on the problems. What I mean by think aloud is to say out loud everything that you say to yourself silently. Just act as if you are alone in the room speaking to yourself.

Participants practised verbalizing while performing three warm-up tasks that were not related to the experiment (Ericsson & Simon, 1993; Fox et al., 2011). All participants received instructions about verbal protocols, and all performed the warm-up tasks. This was carried out in the interest of experimental control, to avoid confounding the inclusion of warm-up tasks before the experiment with the requirement to think aloud during the experiment (Fox et al., 2011).

Following training, participants received instructions about the experiment. Those in the verbalization condition were then told to think aloud throughout the experiment. If a participant was silent for longer than one trial, they were reminded to continue thinking aloud. Participants in the silent condition were told to remain silent.

RESULTS

Outcome and process measures

We began by testing for performance differences between participants in the verbal and silent conditions (i.e., reactivity; Figure 2). The mean difference in response accuracy between conditions was 1.70 with a 95% confidence interval (CI) of (−1.47, 4.88), the difference in pay was .038 ¢ with a 95% CI of (−.79, .72), and the difference in number of cues

DOI: 10.1002/bdm

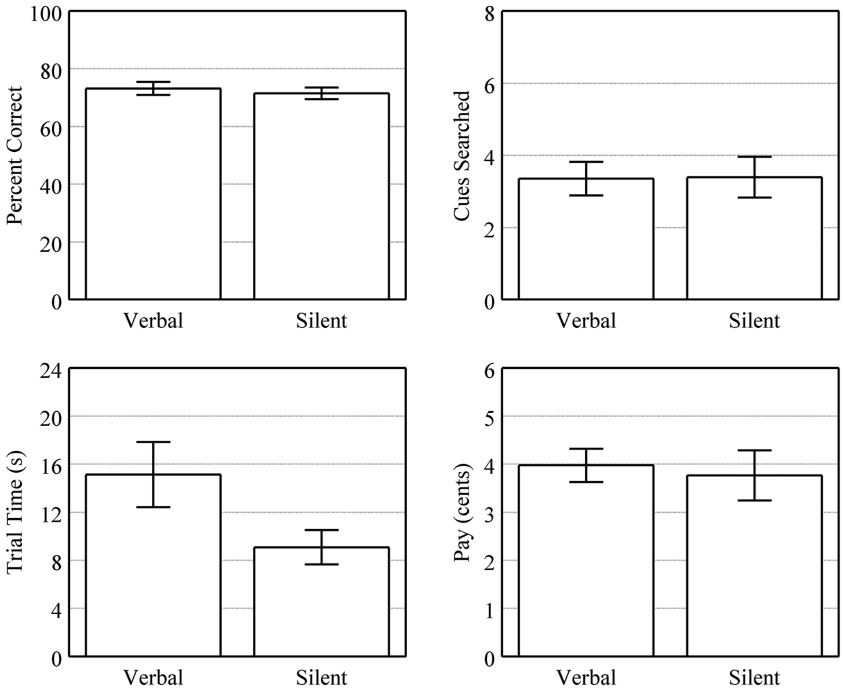

Figure 2. Percent correct (top left), mean number of cues revealed per trial (top right), mean time per trial (bottom left), and mean pay per trial (bottom right). Error bars show 95% confidence intervals for the means

revealed per trial was .21 with a 95% CI of (−.44, .86). None of these differences were statistically significant (accuracy: $t(36) = 1.08$, $p > .1$, Cohen's $d = .35$; cues revealed: $t(36) = .10$, $p > .1$, $d = .03$; pay: $t(36) = .65$, $p > .1$, Cohen's $d = .21$). The difference in time per trial between conditions was 6.05 s with a 95% CI of (2.89, 9.21). This difference was statistically significant, $t(36) = 3.88$, $p < .0001$, Cohen's $d = 1.26$. These results are consistent with the standard finding that verbalization affects the duration of problem-solving operations but not the manner in which problems are solved (Fox et al., 2011).

Next, we examined search behavior (Figure 3). Because cues could be revealed separately for the two stocks, each cue could be selected twice per trial. Participants in both conditions revealed fewer than four cues per trial. A 2 (condition) × 4 (cue validity) mixed-design analysis of variance (ANOVA) revealed a main effect of cue validity,

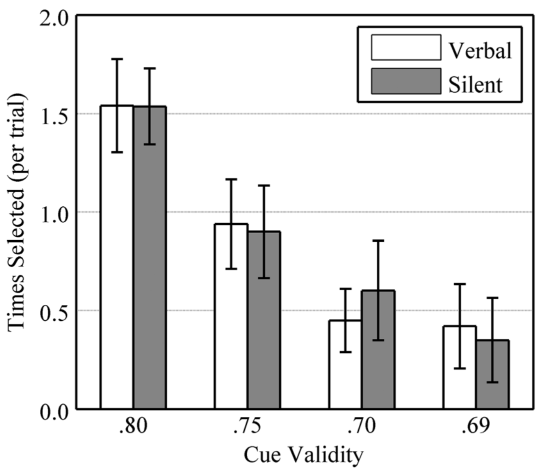

Figure 3. Mean number of times per trial that participants selected each cue (with 95% confidence intervals)

$F(3,108) = 49.92$, $p < .0001$, $\eta^2 = .58$. The main effect of verbalization condition was not significant, $F(1, 36) = .01$, $p > .1$, $\eta^2 = .00$, nor was the interaction, $F(3,108) = .459$, $p > .1$, $\eta^2 = .01$. Participants revealed cues with higher validity more often, and they did so regardless of verbalization condition.

Model-based classification

To characterize the strategies that participants used, we examined the selections predicted by three choice rules: WADD, TAL, and TTB. We used a maximum likelihood approach to calculate the probability of each participant's data separately for the three strategies (Supporting Information). We converted these probabilities into model weights (Liu & Smith, 2009; Wagenmakers, 2007). Weights sum to one across the strategies, with values near one indicating strong support for a strategy and values near zero indicating weak support for a strategy. We then classified each participant as using the single strategy with greatest weight.

Figure 4 shows proportional model weights by participant and condition. For most participants, model weights provided strongest evidence for TTB. For fewer participants, model weights provided strongest evidence for WADD or TAL. Although the model-based analysis strongly discriminated between TTB and the remaining two strategies, it only weakly discriminated between WADD and TAL.

WADD and TAL are compensatory strategies—cues with lower validity can compensate for the value of a cue with higher validity. Conversely, TTB is noncompensatory—cues with lower validity cannot compensate for the value of a cue with higher validity. Because of this distinction and because the model-based analysis did not strongly discriminate between WADD and TAL, we combined these strategies into

DOI: 10.1002/bdm

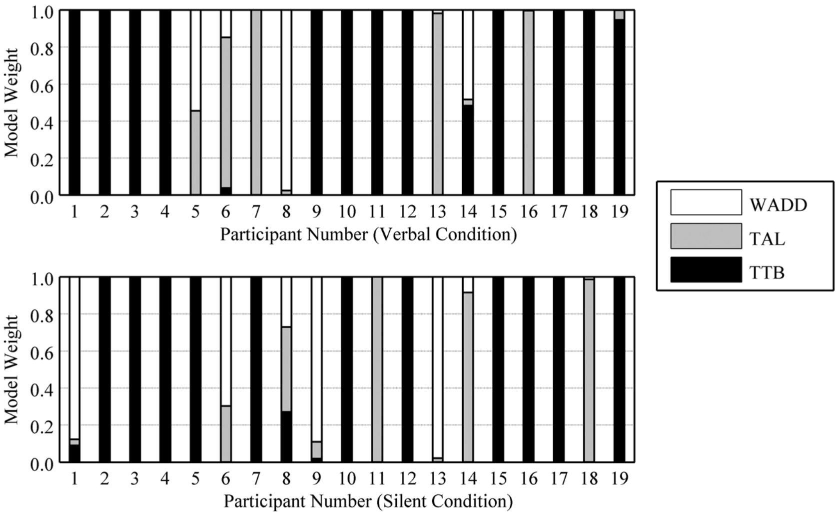

Figure 4. Stacked bar chart showing proportional, likelihood-based model weights for each participant in the verbal and silent conditions. Bars that are predominantly black indicate substantial evidence for take-the-best (TTB), and bars that are predominantly white or gray indicate substantial evidence for weighted additive (WADD) or tally (TAL), respectively

Table 1. Number of participants classified as using compensatory or noncompensatory strategies in verbal and silent conditions

Condition	Compensatory (weighted additive/tally)	Noncompensatory (take-the-best)
Verbal	6	12
Silent	8	11

a single, compensatory group. Table 1 shows the number of participants in each condition whose decisions were best accounted for by a compensatory strategy (WADD or TAL) or a noncompensatory strategy (TTB).[3] A chi-squared test confirmed that these numbers did not differ significantly between conditions, $\chi^2(1) = .30$, $p > .1$.

Next, we examined whether performance measures provided converging evidence for the outcome-based classifications. Figure 5 displays outcome and process measures by condition and strategy classification. A multivariate analysis of variance of the four dependent measures across condition and strategy revealed a significant effect of strategy, $F(3,31) = 7.17$, $p < .001$, $\eta^2 = .41$. We applied a 2 (condition) × 2 (strategy) ANOVA to each measure. Participants who used a noncompensatory strategy were slightly, although not significantly more accurate, $F(1,33) = 2.53$, $p > .1$, $\eta^2 = .07$. They revealed marginally fewer cues, $F(1,33) = 3.92$, $p < .06$, $\eta^2 = .11$, responded more quickly, $F(1,33) = 4.81$, $p < .05$, $\eta^2 = .13$, and earned more money, $F(1,33) = 12.14$, $p < .01$, $\eta^2 = .27$. This indicates a "less-is-more" effect: participants who used TTB revealed fewer cues and were slightly more accurate.

We then examined search behavior by condition and strategy classification (Figure 6). A 2 (condition) × 2 (strategy) × 4 (cue validity) ANOVA revealed main effects of cue validity, $F(3,99) = 50.83$, $p < .0001$, $\eta^2 = .61$, and strategy, $F(1,33) = 3.92$, $p < .05$, $\eta^2 = .11$, and a significant interaction between cue validity and strategy, $F(3,99) = 17.76$, $p < .0001$, $\eta^2 = .35$. Participants who used a noncompensatory strategy revealed cues with higher validity most often, but participants who used a compensatory strategy revealed all cues equally often.

[3]Participant 14 in the verbal condition had identical model weights for WADD and TTB and so could not be clearly assigned one strategy classification. This participant was excluded from subsequent model-based analyses.

Verbal protocols

The experiment yielded about 11 h of verbal data. Protocols were transcribed and segmented into utterances. Utterances were defined as the smallest unit of speech that expressed a complete thought. Their boundaries were denoted by sentence completion, silence, or verbal pauses. In total, participants made 12 602 task-related utterances.

Each utterance was assigned to a category. The coding system was based on prior notions of categories that would be expected in the data given process models of the decision strategies. Broadly, these models specify a sequence of operations corresponding to a search rule, a stopping rule, and a decision rule (Gigerenzer & Gaissmaier, 2011). The coding system was then refined using data from five participants collected during a separate pilot study. The final coding system contained a total of 10 categories and subcategories:

1. Search: express information search.

2. Encoding: reading the value of one indicator for one stock.

3a. Single-indicator elaboration: comparing one indicator between stocks.

3b. Multi-indicator elaboration: aggregating across multiple indicators within a stock or comparing multiple indicators between stocks.

4a. Unjustified decision: stating selection without justification.

DOI: 10.1002/bdm

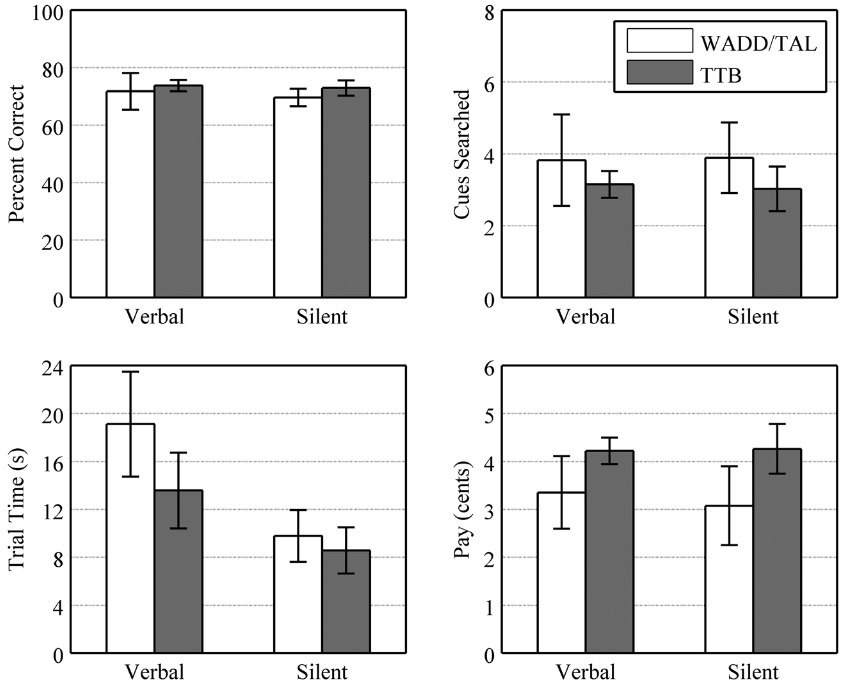

Figure 5. Percent correct (top left), mean number of cues revealed per trial (top right), mean time per trial (bottom left), and mean pay per trial (bottom right) (with 95% confidence intervals). Bar colors correspond to participants classified as using compensatory (WADD/TAL) or noncompensatory (TTB) strategies in each experiment condition

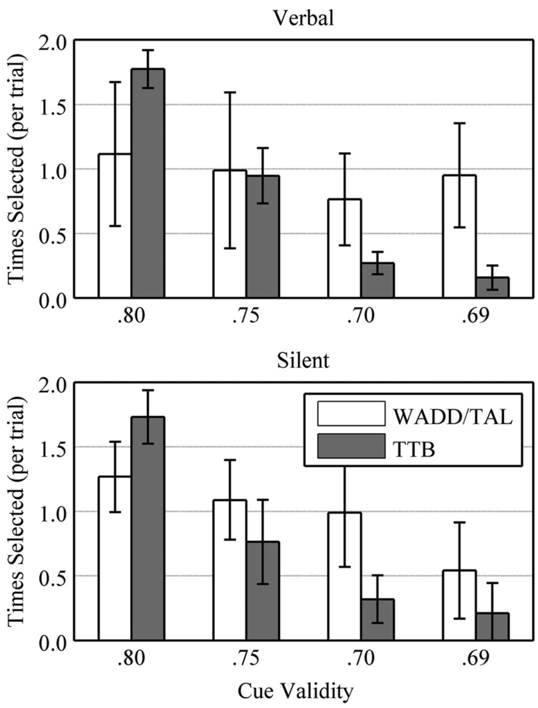

Figure 6. Mean number of times per trial that participants revealed each cue (with 95% confidence intervals). Bar colors correspond to participants classified as using compensatory (WADD/TAL) or noncompensatory (TTB) strategies in each experiment condition

4b. Single-indicator decision: justification with a single indicator.

4c. Multi-indicator decision: justification with multiple indicators.

5. Evaluation: reaction to trial outcome.

6. Metacognitive: general statement about task structure or strategy.

7. Uncategorized: statement that does not fit within existing categories.

Table 2 contains actual statements from each of these, excluding uncategorized statements.

One investigator coded 100% of the utterances, and a second investigator coded 10% of the utterances drawn randomly from each participant. To increase accuracy, coding was performed with the aid of audiovisual recordings of trials. Less than 1% of protocols were left uncategorized. The numbers of utterances assigned to the nine substantive categories by each investigator were nearly identical, $\chi^2(8) = 1.09$, $p > .1$. The mean inter-rater reliability for individual participants, measured by Cohen's kappa, was .97 with a 95% CI of (.90, .98).

After entering utterances into the primary coding system, we identified all statements that contained information about an indicator's validity. Additionally, we identified all statements that contained information about search costs, losses, and trial pay. We entered these a second time into a separate coding system as validity statements and monetary statements. Table 3 contains actual statements from these categories. Inter-rater reliability for validity statements was perfect (1.0), and inter-rater reliability for monetary statements was near perfect (.96).

Results

Figure 7 shows the overall number of statements assigned to the six primary categories separately for participants

DOI: 10.1002/bdm

Table 2. Coding categories with example statements

Category	Statement
1. Search	So, let us do share trend positive. Financial reserves next.
2. Encoding	So, that one is no. TJG is yes for share trend.
3a. Single-indicator elaboration	Both are yes for financial reserves. Only BQN invests in new projects.
3b. Multi-indicator elaboration	I have two yesses and two noes for QSV. Each company has a yes and a no.
4a. Unjustified decision	Choose JBL. I will guess the one on the left.
4b. Single-indicator decision	Pick the one who has reserves. I will go for the one that has yes for the top most.
4c. Multi-indicator decision	With three yesses and two noes, I will go with HBP. Pick the two yesses.
5. Evaluation	Earned 6 ¢. Dangit!
6. Metacognitive	Contemplating just trying to see based on the first indicator. So, I do not know if these things actually follow market trends.

Table 3. Validity and monetary categories with example statements

Category	Statement
Validity	The next highest is 75%. Both are negative for the most valid indicator.
Monetary	So, rather than spending a cent, I am going to choose MVG. Earned 6 ¢ in that trial.

classified as using noncompensatory or compensatory strategies. A 2 (strategy) × 6 (category) mixed-measures ANOVA revealed main effects of strategy, $F(1,16) = 6.53$, $p < .05$, $\eta^2 = .29$, and category, $F(5,80) = 31.57$, $p < .001$, $\eta^2 = .66$, and a significant interaction, $F(5,80) = 4.92$, $p < .01$, $\eta^2 = .24$. Participants who used a compensatory strategy made more utterances. Specifically, they made more search statements, $t(16) = 2.44$, $p < .05$, Cohen's $d = 1.22$, encoding statements, $t(16) = 2.35$, $p < .05$, Cohen's $d = 1.17$, and evaluation statements, $t(16) = 3.21$, $p < .01$, Cohen's $d = 1.61$. The increased number of search and encoding statements is consistent with the fact that these participants revealed more indicators.

The elaboration and decision subcategories are especially informative with respect to the question of strategy use. Noncompensatory strategies like TTB are associated with indicator-based processing. Consequently, participants using a noncompensatory strategy should make more single-indicator elaboration and decision statements. Alternatively, compensatory strategies like WADD and TAL are associated with alternative-based processing. Consequently, participants using compensatory strategies should make more multi-indicator elaboration and decision statements. To test these hypotheses, we calculated the relative proportions of single-indicator and multi-indicator statements in the elaboration and decision categories (Table 4), that is, the number of single-indicator statements divided by the combined number of single-indicator and multi-indicator statements. A 2 (strategy) × 2 (category) mixed-measures ANOVA revealed a significant effect of strategy, $F(1,16) = 20.20$, $p < .0001$, $\eta^2 = .57$. Participants who used TTB made more single-indicator elaboration statements, $t(16) = 3.12$, $p < .01$, Cohen's $d = 1.56$, and decision statements, $t(16) = 4.39$, $p < .001$, Cohen's $d = 2.19$, as one would expect given the noncompensatory nature of that strategy.

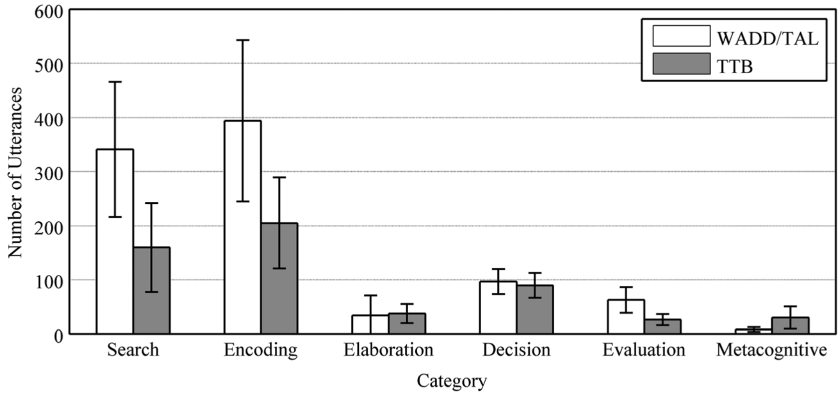

Figure 7. Mean number of utterances in six categories and for participants classified as using compensatory (WADD/TAL) or noncompensatory (TTB) strategies (with 95% confidence intervals)

DOI: 10.1002/bdm

Table 4. Relative proportion of single-indicator elaboration and decision statements for participants classified as using compensatory (WADD/TAL) or noncompensatory (TTB) strategies (with 95% confidence intervals)

Category	WADD/TAL	TTB
Single-indicator elaboration	.44 (.26, .62)	.83 (.69, .97)
Single-indicator decision	.29 (.07, .51)	.85 (.72, .98)

We then analyzed the numbers of validity and monetary statements. On average, participants made a total of 36 validity statements with a 95% CI of (0, 79) and 14 monetary statements with a 95% CI of (6, 22) during the experiment. Because the validity and monetary statement counts were not normally distributed, we used the nonparametric Wilcoxon rank-sum test to compare the median numbers of statements between strategy subgroups. The number of validity statements did not differ between participants who used compensatory and noncompensatory strategies (11.0 vs 13.5, $p > .1$). However, participants who used a compensatory strategy made fewer monetary statements (3.0 vs 12.5, $p < .05$).

As part of a final, exploratory analysis, we reviewed participants' metacognitive statements. Because of their diversity, metacognitive statements could not easily be divided among subcategories. However, a small number of metacognitive statements contained clear descriptions of decision processes.[4] Table 5 contains examples from 10 participants. Some of the descriptions express TTB's search rule—examine cues in order of their validity (Participants 1, 3, and 12). Others express TTB's stopping rule—terminate search upon finding a discriminating cue (Participants 3, 9, 15, 17, and 18). Still, others express TTB's decision rule—choose the alternative with the positive cue value (Participants 9, 11, 17, and 18). In contrast to these reports, some statements indicate compensatory processing and tallying (Participants 5 and 14). There were too few descriptions of the decision process to perform statistical analyses, and some participants never provided descriptive metacognitive statements. However, of the descriptive statements that were provided, most were consistent with the model-based classification of participants' strategies, which are also contained in Table 5.

Multidimensional classification

In the preceding sections, we used a model-based analysis to classify participants based on their decisions. We then analyzed search behavior and verbal reports as a function of strategy classification. This is the predominant approach used in the literature. A potentially more powerful way to classify participants is based jointly on outcome measures, search measures, and verbal measures—that is, a multimethod approach (Schulte-Mecklenbeck et al., 2011). To the extent that measures are related but distinct (Schulte-Mecklenbeck, Sohn, de Bellis, Martin, & Hertwig, 2013), the inclusion of multiple criteria enhances strategy classification.

To combine information across the set of outcome and process measures, we performed a multidimensional classification. For each participant, we recorded the model weight assigned to TTB (the noncompensatory strategy), the average number of cues revealed per trial, and the combined proportion of single-indicator elaboration and decision statements. To place equal emphasis on each of these dimensions, we normalized the three dependent variables using z-scores. We then used a two-step k-means cluster analysis to determine the number of clusters that minimized the Bayesian information criterion and to assign each participant to a cluster.

The analysis revealed two clusters of 10 and nine participants (Figure 8). The first cluster had large weight assigned to noncompensatory decisions, searched few cues, and had many single-indicator statements. This constellation of measures is most consistent with TTB. The second cluster had low weight assigned to noncompensatory decisions, tended to search many cues, and had many multi-indicator statements. This constellation of measures is most consistent with WADD or TAL. The results of the multidimensional classification were nearly identical to the results of the model-based analysis, with the exception of two participants. Both made decisions consistent with TTB. However, both also searched many cues and made many multi-indicator statements. Classifications based only on outcome measures obscure these internal inconsistencies.

Which of the three dimensions is most diagnostic? Search behavior was only weakly correlated with decisions and verbal reports ($r = .30$ and .55, respectively), whereas decisions and verbal reports were strongly correlated ($r = .73$). In other words, decisions and verbal reports were most internally consistent.

Separation between clusters varied along the three dimensions. Figure 9 shows the distributions of values for search, decisions, and verbal reports. Each panel collapses data across two of the dimensions from Figure 8 and shows the distribution of participants' values along the remaining dimension. The distributions for individuals assigned to different clusters overlap completely for the search measure but are almost nonoverlapping for the decision and verbal measures. Sensitivity, defined as the difference between the clusters' means divided by their pooled standard deviations, was low for search behavior ($d' = .97$), moderate for decisions ($d' = 2.38$), and high for verbal reports ($d' = 4.95$). These analyses suggest that outcome measures and verbal reports are more useful than a particular search measure, information acquisition, for inferring participants' decision strategies in multiple-cue probabilistic inference.

We classified participants based on three dependent variables all thought to arise from the underlying strategy models. What else beyond these dependent measures differentiates participants in the two clusters from one another? Individual differences are ubiquitous in studies of strategy selection (Scheibehenne, Rieskamp, & Wagenmakers, 2013). Factors such as age, intelligence, and working memory capacity are thought to influence people's tendency to adopt heuristics like TTB (Bröder, 2011). However, further research is needed to understand why people adopt the different strategies that they do.

[4]Participants were not explicitly told to describe their strategies during the task. Accordingly, few statements contained such descriptions (<.05%).

DOI: 10.1002/bdm

Table 5. Example metacognitive statements

Identification	Outcome-based classification	Statement
1	TTB	I am always going to start with 80% because I think that would be the best indicator, because it is the highest percent.
3	TTB	I think that I am going to stay with the method of starting with the 80, going to 75, and then just keep going until one is yes and one is no.
5	WADD / TAL	So, I think that this time, I am going to check more indicators than last time because last time I had two yesses and I thought it was a solid choice, but it was not the right choice.
9	TTB	Look at established company, and if they are both the same, see if they invest in new projects. I am just going to go with all the ones that say yes for established company compared with a no in the other company.
11	TTB	Every time one is yes and the other is no, I always went with that one.
12	TTB	I am just going to go with financial reserves because that is the highest.
14	WADD / TAL	And if it is two or three typically, that is what you are looking for. Most do not have all of them.
15	TTB	I decided not to uncover the fourth one because the last time I did and they had two different answers, I ended up going with established company anyway. So, there is really no information that would sway my vote.
17	TTB	Find the share trend, and then if the share trend of one then the other, then if they are both the same, then look at established company. Whichever is yes first, then choose it.
18	TTB	I should just guess one category, and if I obtain just a single yes, I should jump on that. But if I acquire matching answers, then I should reveal, continue revealing categories. That might be the trick here.

WADD, weighted additive; TAL, tally; TTB, take-the-best.
Outcome-based classification was determined in advance based on participants' decisions.

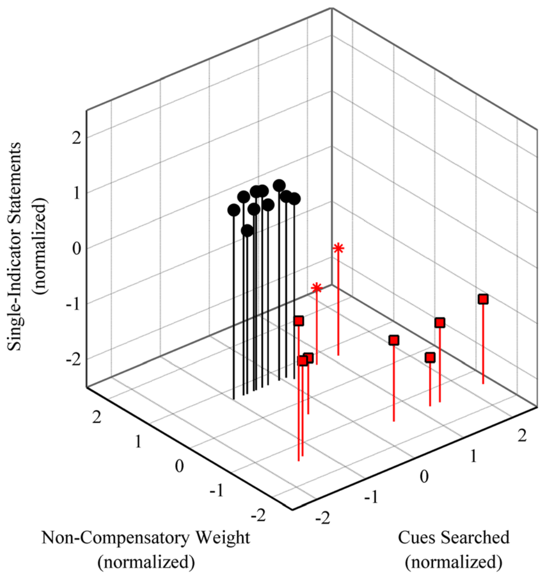

Figure 8. Scatter plot of normalized search measure (cues searched), outcome measure (noncompensatory decision weight), and verbal measure (proportion of single-indicator elaboration and decision statements). Black circles denote participants whose behavior is consistent with take-the-best, and red squares denote participants whose behavior is consistent with weighted additive or tally. Stars denote two participants classified differently by multidimensional and outcome-based approaches

GENERAL DISCUSSION

The primary goal of this research was to understand the strategies that people use to perform multiple-cue probabilistic inference. The experiment revealed four important results. First, the decisions of the majority of participants were consistent with the noncompensatory TTB strategy. That said that the decisions of a substantial minority were consistent with compensatory strategies such as WADD or TAL. Second, participants who decided according to TTB also revealed the fewest indicators, providing additional support for the outcome-based classifications. Third, these same participants made more elaboration statements that compared a single indicator between stocks, and they justified more decisions based on a single indicator. The convergence of evidence across outcome measures, search measures, and verbal reports strongly supports the conclusion that many, although not all, participants adopt a TTB decision process in this kind of environment. Fourth, among the dependent measures assessed in this study, verbal reports showed the highest sensitivity, demonstrating their value as a source of cognitive process data.

Classic theories from judgment and decision making research specify the relationship between the input of the decision process (i.e., the stimulus configuration) and its output (i.e., the decision). In addition to accounting for outcomes, the heuristics models that we evaluated also seek to explain the sequence of psychological processes that accompany decisions. To rigorously test these models, one must analyze their predictions concerning both outcome measures and process measures (Schulte-Mecklenbeck et al., 2011). A priori, it was unknown whether verbal reports, a high-density process measure, would be consistent with the predictions of TTB. We found that they were in participants whose decisions were predicted by TTB. Thus, this study provides novel empirical evidence for TTB as a process model of multiple-cue probabilistic inference.

These results are consistent with earlier studies that found evidence of TTB based on outcome and search measures alone (Bröder, 2011; Rieskamp & Otto, 2006). The verbal data provide more than just marginally stronger support for

DOI: 10.1002/bdm

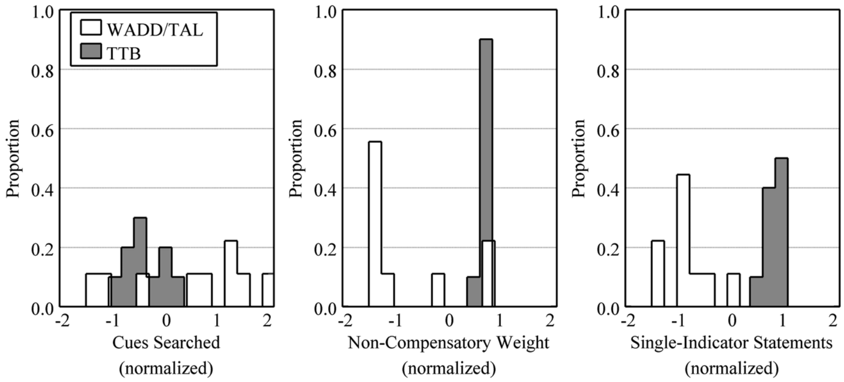

Figure 9. Distributions of values for search measure, outcome measure, and verbal measure. Gray distribution corresponds to individuals assigned to the take-the-best (TTB) cluster, and white distribution corresponds to individuals assigned to the weighted additive (WADD) and tally (TAL) cluster

TTB, however. Verbal reports were more diagnostic with respect to individual differences than search measures. Search and stopping rules are distinct from decision rules. Participants may acquire information without using it (Bröder, 2003; Harte & Koele, 2001; Maule, 1994; Newell & Shanks, 2003; Reisen et al., 2008). Verbal protocols revealed how participants used the information that they acquired to make decisions. Verbal reports were also more diagnostic than outcome measures. Although outcome measures provide information about decision rules, different strategies often predict the same choices. Verbal protocols discriminated among strategies even when they predicted the same outcomes. For example, in this dataset, verbal reports provided information about how each participant made selections during about 30% of trials where both strategies predicted the same choice. This more than doubles the number of classifiable trials.

In addition to enhancing diagnostic power, the verbal data suggested one reason that some participants adopted more frugal strategies. Participants who used TTB showed greater awareness of the monetary cost of information acquisition and the payoff structure of the task. Although this propensity could be inferred from search behavior and choices, participants' verbal reports contained direct evidence that they considered these factors. In sum, the verbal data provide information about which strategies participants use and why.

The second goal of this research was to determine whether verbalization influences the cognitive processes involved in multiple-cue probabilistic inference. Verbalization did not affect the percentage of correct responses, the number of indicators revealed, or which indicators were revealed. Verbalization also did not affect the proportion of individuals classified as using noncompensatory or compensatory strategies. Participants in the verbal condition did take longer to make decisions, however. These results are consistent with the view that verbalization affects the duration of problem-solving operations but not the manner in which problems are solved (Fox et al., 2011).

Adaptive decision making

According to the adaptive toolbox theory of cognition, the mind contains a collection of decision heuristics, one of which is TTB (Gigerenzer & Gaissmaier, 2011; Gigerenzer et al., 1999). TTB is not a universal theory of probabilistic inference—it is not intended to account for the choices of all people at all times. The ecological rationality of TTB and other heuristics lies in their match to the structure of the environment (Gigerenzer & Goldstein, 1996; Hogarth & Karelaia, 2007; Simon, 1955). The adaptive toolbox theory predicts that people will only use TTB when it is adaptive to do so. For example, when cue validities are noncompensatory, when information acquisition is costly, and when time pressure is great.

Most of these conditions were met in our experiment. A small number of cues had high validity. Additionally, cue values were concealed, and information acquisition was costly in terms of time, physical effort, and money. Accordingly, more than half of participants appeared to use TTB. This proportion is comparable with other experiments with environment structures that favor noncompensatory strategies (Bröder, 2011).

The adaptive toolbox theory predicts that people will use other strategies like WADD and TAL when cues have moderate-to-low validity, when information is freely available, and when time pressure is minimal. Empirical studies have confirmed these predictions (Bröder, 2000, 2003; Payne et al., 1988; Rieskamp & Otto, 2006). If we replicated our experiment in such conditions, we expect that more participants would use compensatory strategies like WADD and TAL. More of their elaboration statements would involve combining information across multiple indicators within a stock, and more of their decision statements would refer to multiple indicators. Although these predictions remain to be tested, participants classified as using compensatory strategies in the current experiment did exhibit the expected verbalization patterns for elaboration and decision statements.

LIMITATIONS

Despite its widespread use, verbal protocol analysis remains controversial because of the concern that verbalization affects the cognitive processes being investigated (Schooler, 2011). For example, Russo, Johnson, and Stephens (1989) found that concurrent verbal reports increased accuracy when participants were asked to choose between pairs of gambles

DOI: 10.1002/bdm

and decreased accuracy when they solved addition problems. Enhanced performance is thought to arise from elaborative processing caused by the instruction to verbalize thoughts, and impaired performance is thought to arise from the cognitive demands associated with verbalizing concurrently.

Ericsson and Simon (1993) distinguished between three levels of verbalization: level 1 involves reporting verbal thoughts from short-term memory, level 2 involves translating and reporting nonverbal thoughts from short-term memory, and level 3 involves interpreting thoughts and mental processes (i.e., introspection). Only level 3 verbalizations increase the potential for reactivity (Ericsson & Simon, 1993; Fox et al., 2011). Our instructions specifically encouraged level 1 verbalizations. Accordingly, comparison with the requisite, silent control condition revealed little evidence of reactivity.

Another criticism of verbal reports is that they are incomplete and inaccurate (Nisbett & Wilson, 1977; Russo et al., 1989; Schooler, 2011). In fact, the model by Ericsson and Simon (1993) assumes that verbalizations will be incomplete. Participants are asked to verbalize information in the current focus of attention and working memory, which constitutes only part of the ongoing mental processes. To overcome this limitation of protocol analysis, we gathered additional outcome and process measures concurrently with verbal reports. Doing so also allowed us to examine the veracity of participants' reports. We observed substantial consistency across decisions, search behavior, and verbalizations.

A different concern pertains to the presentation format used in our experiment. Cue values were concealed within a grid. To view a cue value, participants used a cursor to select the corresponding cell in the grid. This presentation format is based on the popular Mouselab computerized information board paradigm (Payne et al., 1988; for reviews, see Bröder, 2011; Harte et al., 1994). However, some researchers have expressed concern with this technique. Lohse and Johnson (1996) studied search behavior using Mouselab and eye tracking. They found that participants attended to more information when it was visible versus when it was concealed. Likewise, Glöckner and Betsch (2008) argued that information search demands, rather than information integration demands, cause people to adopt noncompensatory strategies in Mouselab experiments.

We have several reactions to these concerns. First, some studies report only minor differences when comparing Mouselab and eye tracking with respect to their impact on search and decision behavior (Reisen et al., 2008). Second, even when information is visible, processing must occur sequentially through a series of saccades and fixations (Orquin & Mueller Loose, 2013). Experiments in which cue values are visible or concealed simply impose different information acquisition costs. Third, people are sensitive to strategy costs on the millisecond level (Gray, Sims, Fu, & Schoelles, 2006), and the adaptive toolbox theory predicts that they will select heuristics suitable for the structure of the environment. If we presented information openly, we expect that more participants would use WADD or TAL. Gathering verbal reports in such a context would provide especially useful information about information search and integration. Finally, many real-world decisions impose costly, time-consuming information search (e.g., a medical professional ordering diagnostic tests). Mouselab-style experiments are informative with respect to these types of decisions.

Two final limitations warrant mention. First, the sample sizes in each condition of the experiment were relatively small. Increasing the number of participants would increase the statistical power of comparisons between the silent and verbal conditions. However, given the small effect sizes, the results would not likely change. Second, this study deals with a single variant of multiple-cue probabilistic inference: selection between a pair of alternatives based on cues with binary values. Future studies should examine selection among three or more alternatives based on cues with continuous values.

CONCLUSION

Simon (1992) stated that our methods for gathering data must fit the shapes of our theories; the proper tool must be selected for the job (Gluck, Staszewski, Richman, Simon, & Delahanty, 2001). The primary innovation of this work is the application of verbal protocol analysis to the problem of multiple-cue probabilistic inference. Participants' verbal reports provided critical evidence about how they used information that they acquired, evidence that standard process and outcome measures do not adequately capture. Namely, participants tended to compare individual indicators between stocks, and they based most decisions on one indicator. This style of information integration is consistent with the use of a noncompensatory decision strategy such as TTB. Of course, people do not always TTB. The method used here, verbal protocol analysis, has considerable potential to enhance understanding of the other diverse strategies that people adopt in different environments and for different types of problems.

ACKNOWLEDGEMENTS

This research was performed while Matthew Walsh held a National Research Council Research Associateship Award with the Air Force Research Laboratory's (AFRL) Cognitive Models and Agents Branch. The authors thank AFRL for supporting this research, and they thank members of the Cognitive Models and Agents Branch for stimulating conversation and critical thinking on this topic. Special thanks to Michael Collins for transcribing the verbalizations. Rachel Boyd, Michael Collins, Michael Schulte-Mecklenbeck, Thora Tenbrink, and two anonymous reviewers provided constructive comments on earlier versions of this paper.

REFERENCES

Ayal, S., & Hochman, G. (2009). Ignorance or integration: The cognitive processes underlying choice behavior. *Journal of Behavioral Decision Making*, *22*, 455–474.

DOI: 10.1002/bdm

Brandstätter, E., & Gussmack, M. (2013). The cognitive processes underlying risky choice. *Journal of Behavioral Decision Making*, *26*, 185–197.

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. (2011). The quest for take the best – insights and outlooks from experimental research. In Todd, P., Gigerenzer, G., & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world* (p. 364–380). New York, NY: Oxford University Press.

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

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90*, 63–86.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*, 20–33.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In Gigerenzer, G., Todd, P. M., & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (p. 97–118). New York, NY: Oxford University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*, 199–213.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380–417.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171–192.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal protocols as data*. Cambridge, MA: MIT Press.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*, 316–344.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

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

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

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

Glöckner, A., & Betsch, T. (2010). Accounting for critical evidence while being precise and avoiding the strategy selection problem in a parallel constraint satisfaction approach: A reply to Marewski (2010). *Journal of Behavioral Decision Making*, *23*, 468–472.

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, *139*, 532–542.

Gluck, K. A., Staszewski, J. J., Richman, H., Simon, H. A., & Delahanty, P. (2001). The right tool for the job: Information-processing analysis in categorization. Proceedings of the 23rd Annual Meeting of the Cognitive Science Society (pp. 330–335). London: Erlbaum.

Gray, W. D., Sims, C. R., Fu, W. T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*, 461–482.

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

Harte, J. M., Westenberg, M. R. M., & van Someren, M. (1994). Process models of decision making. *Acta Psychologica*, *87*, 95–120.

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

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*, 733–758.

Johnson, E. J., & Meyer, R. J. (1984). Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research*, *11*, 528–541.

Johnson, E. J., & Schkade, D. A. (1989). Bias in utility assessments: Further evidence and explanations. *Management Science*, *35*, 406–424.

Katsikopoulos, K. V., Schooler, L. J., & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*, 1259–1266.

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.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

Liu, C. C., & Smith, P. L. (2009). Comparing time–accuracy curves: Beyond goodness-of-fit measures. *Psychonomic Bulletin & Review*, *16*, 190–203.

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, *68*, 28–43.

Martignon, L., & Hoffrage, U. (1999). *Why does one reason decision making work? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), Simple heuristics that make us smart (p. 119–140)*. New York, NY: Oxford University Press.

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*, 796–810.

Maule, A. J. (1994). A componential investigation of the relation between structural modelling and cognitive accounts of human judgement. *Acta Psychologica*, *87*, 199–216.

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Science*, *9*, 11–15.

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

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*, 190–206.

Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers? *Acta Psychologica*, *144*, 97–103.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol

DOI: 10.1002/bdm

analysis. *Organizational Behavior and Human Performance*, *16*, 366–387.
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.
Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1088–1104.
Ranyard, R., & Svenson, O. (2011). Verbal data and decision process analysis. In M. Schulte-Mecklenbeck, A. Kuhberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide* (p. 115–139). New York, NY: Taylor & Francis.
Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making*, *3*, 641–658.
Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*, 258–276.
Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236.
Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*, 759–769.
Scheibehenne, B., Rieskamp, J., & Wagenmakers, E. J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*, 39–64.
Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, *137*, 345–350.
Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, *6*, 733–739.
Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, *71*, 242–251.
Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*, 207–222.
Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.
Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*, 150–161.
Söllner, A., Bröder, A., Glöckner, A., & Betsch, T. (2014). Single-process versus multiple-strategy models of decision making: Evidence from an information intrusion paradigm. *Acta Psychologica*, *146*, 84–96.
Tenbrink, T., & Wiener, J. (2009). The verbalization of multiple strategies in a variant of the traveling salesperson problem. *Cognitive Processing*, *10*, 143–161.
Timmermans, D. (1993). The impact of task complexity on information use in multi-attribute decision making. *Journal of Behavioral Decision Making*, *6*, 95–111.
von Neumann, J., & Morgenstern, O. (1944). *Theory of gambles and economic behavior*. Princeton, NJ: Princeton University Press.
Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.
Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53–85.

Authors' biographies:

Matthew M. Walsh is a contractor for the Air Force Research Laboratory's Cognitive Models and Agents Branch. Previously, he earned his PhD in psychology from Carnegie Mellon University. He uses behavioral and neuroimaging techniques and computational modeling to study human decision making.

Kevin A. Gluck is a Principal Cognitive Scientist with the Air Force Research Laboratory. His interests are in the areas of learning, memory, and decision making, with applications in education, training, and autonomous systems. He uses empirical research and cognitive modeling to study these topics.

Authors' addresses:

Matthew M. Walsh, Air Force Research Laboratory Wright-Patterson Air Force Base, OH, USA

Kevin A. Gluck, Air Force Research Laboratory Wright-Patterson Air Force Base, OH, USA

DOI: 10.1002/bdm