

Satisficing in Split-Second Decision Making Is Characterized by Strategic Cue Discounting

Hanna Oh and Jeffrey M. Beck
Duke University

Pingping Zhu
Cornell University

Marc A. Sommer
Duke University

Silvia Ferrari
Cornell University

Tobias Egner
Duke University

Much of our real-life decision making is bounded by uncertain information, limitations in cognitive resources, and a lack of time to allocate to the decision process. It is thought that humans overcome these limitations through *satisficing*, fast but “good-enough” heuristic decision making that prioritizes some sources of information (cues) while ignoring others. However, the decision-making strategies we adopt under uncertainty and time pressure, for example during emergencies that demand split-second choices, are presently unknown. To characterize these decision strategies quantitatively, the present study examined how people solve a novel multicue probabilistic classification task under varying time pressure, by tracking shifts in decision strategies using variational Bayesian inference. We found that under low time pressure, participants correctly weighted and integrated all available cues to arrive at near-optimal decisions. With increasingly demanding, subsecond time pressures, however, participants systematically discounted a subset of the cue information by dropping the least informative cue(s) from their decision making process. Thus, the human cognitive apparatus copes with uncertainty and severe time pressure by adopting a “drop-the-worst” cue decision making strategy that minimizes cognitive time and effort investment while preserving the consideration of the most diagnostic cue information, thus maintaining “good-enough” accuracy. This advance in our understanding of satisficing strategies could form the basis of predicting human choices in high time pressure scenarios.

Keywords: decision making, satisficing, bounded rationality, cue integration, time pressure

The study of rational decision-making has traditionally focused on decision strategies that maximize utility, referred to as *substantive* (or *unbounded*) *rationality* (Simon, 1955, 1990). These strategies involve exhaustive computations based on perfect knowl-

edge of decision-relevant information, possible choices, and their outcome probabilities and consequences. In the real (large) world, however, rationality is usually *bounded* by incomplete knowledge of the decision-relevant information and limitations placed upon cognitive resources and available computation time, rendering optimal decision-making nearly impossible (Simon, 1955, 1956, 1990). Humans are thought to overcome these limitations via *satisficing*, the use of heuristic strategies that simplify decision-making problems by prioritizing some sources of information while ignoring others (Simon, 1955, 1956, 1990; see Gigerenzer & Gaissmaier, 2011 for a review). This approach leads to solutions that are not precisely optimal but suffice to satisfy some specified criterion level, thereby facilitating fast and “good-enough” decision-making. Even though this type of bounded decision-making likely underpins most decisions we make in everyday life, the manner in which satisficing is triggered and accomplished remains poorly understood. Here, we characterize some of the principles that govern satisficing decision-making due to uncertain information and high time pressure, by combining a novel probabilistic classification task with recently developed Bayesian strategy analysis techniques.

The most basic heuristic shortcut to decision making is memory: When faced with familiar decision problems or choices, people

This article was published Online First June 2, 2016.

Hanna Oh, Department of Psychology and Neuroscience and Center for Cognitive Neuroscience, Duke University; Jeffrey M. Beck, Department of Neurobiology and Center for Cognitive Neuroscience, Duke University; Pingping Zhu, Department of Mechanical and Aerospace Engineering, Cornell University; Marc A. Sommer, Department of Biomedical Engineering, Center for Cognitive Neuroscience, and Duke Institute for Brain Sciences, Duke University; Silvia Ferrari, Department of Mechanical and Aerospace Engineering, Cornell University; Tobias Egner, Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, and Duke Institute for Brain Sciences, Duke University.

This work was funded by Office of Naval Research Grant ONR N000141310561.

Correspondence concerning this article should be addressed to Hanna Oh or Tobias Egner, Department of Psychology and Neuroscience and Center for Cognitive Neuroscience, Duke University, Levine Science Research Center, Box 90999, Durham, NC 27708. E-mail: hanna.oh@duke.edu or tobias.egner@duke.edu

often make memory-based decisions using strategies such as the recognition heuristic (Goldstein & Gigerenzer, 2002; see also Newell & Shanks, 2004; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011; Pohl, 2006), the fluency heuristic (Jacoby & Brooks, 1984; Schooler & Hertwig, 2005; Whittlesea, 1993), and the exemplar-based approach (Juslin, Olsson, & Olsson, 2003; Juslin & Persson, 2002; Nosofsky & Palmeri, 1997). By contrast, when dealing with unfamiliar situations, people appear to take into account the informational structure of the decision problem, such as the relative values and intercorrelations between different sources of information and the cost of acquiring new information, in adopting decision heuristics (Bröder, 2000, 2003; Dieckmann & Rieskamp, 2007; Newell & Shanks, 2003; Rieskamp & Otto, 2006). For example, one way to simplify a decision problem is to evaluate one piece of information (i.e., one cue) at a time, starting from what is considered the most important to less important cues, based on a subjective rank order. In case of a binary choice problem, one well-known lexicographic strategy is the “take-the-best” heuristic, which sequentially searches through cues in descending order of their assumed values and stops upon finding the first (highest-ranked) cue that discriminates between the two alternatives (Gigerenzer & Goldstein, 1996). A choice is made without further evaluating less valuable available cues, thereby simplifying the decision problem. Therefore, in contrast to compensatory strategies that weight and integrate all decision-relevant information, take-the-best is noncompensatory because a deterministic cue cannot be outweighed by any combination of less valuable cues (Gigerenzer & Goldstein, 1996).

Laboratory studies that have documented participants' spontaneous use of this cue-based satisficing strategy have typically provided subjects with a set of cues of varying, explicitly known values, and then assessed the manner in which subjects combine the different cues as a function of satisficing pressures (e.g., redundant cue information, information costs; e.g., Bröder, 2003; Dieckmann & Rieskamp, 2007; Newell, Weston, & Shanks, 2003; Payne, Bettman, & Johnson, 1988; Rieskamp, 2006; Rieskamp & Otto, 2006). In addition, a majority of these studies adopted a serial information search paradigm (Payne et al., 1988), which enables easy tracking of choice patterns but also constrains behavior by hindering quick comparisons of multiple pieces of information (Glöckner & Betsch, 2008). Studies that did not employ this type of paradigm have instead often adopted an outcome-oriented approach, in which certain choices are mapped on to certain decision models (e.g., Bergert & Nosofsky, 2007; Lee & Cummins, 2004). This provides a convenient way to infer strategies but limits the number of trial types that could be used as probes. Additionally, to facilitate the learning of multiple cue values, many studies have provided participants with trial-by-trial deterministic feedback indicating whether they made a correct or incorrect judgment (e.g., Bergert & Nosofsky, 2007; Juslin et al., 2003; Lamberts, 1995; Lee & Cummins, 2004; Pachur & Olsson, 2012; Rieskamp & Otto, 2006). While this work has produced valuable insights into adaptive shifts in decision strategies, it arguably falls short of simulating satisficing in the real world, where we are often exposed to multiple sources of information simultaneously and must infer their approximate values from experience. For example, in deciding whether and when to merge into an exit lane on a busy highway, we must estimate our own speed, the distance to the exit ramp, and the relative distances and speeds of cars traveling in

front of us, behind us, and in the lane we would like to merge into, as well as their drivers' intentions (e.g., indicators). Estimates of all of these cues are uncertain, and their relative importance for avoiding an accident is inferred from previous experience entailing probabilistic feedback (e.g., we do not get in an accident every time we ignore the rear mirror).

The presence of time pressure is known to influence the decision process (for an overview, see Svenson & Maule, 1993), fostering the use of heuristics that can be applied quickly within a choice deadline. Several studies have shown that, under time pressure, people engage in a more selective information search (Böckenholt & Kroeger, 1993; Lamberts, 1995; Maule, 1994; Payne et al., 1988; Rieskamp & Hoffrage, 2008) and employ memory-based heuristics when possible (Goldstein & Gigerenzer, 2002; Lamberts, 1995, 2000; Nosofsky & Palmeri, 1997; Pachur & Hertwig, 2006). In fact, under certain circumstances, decisions made under high time constraints or limited cue exposure can even be found to be more accurate compared to those made after a long deliberation period (Ballew & Todorov, 2007; Wilson & Schooler, 1991), suggesting that the use of fast heuristics can sometimes lead to better choices (see also Gigerenzer & Gaissmaier, 2011). In the perceptual decision-making literature, making choices under time constraints often results in a speed-accuracy trade-off that is characterized by a (conservative) shift of the decision or response threshold (see, e.g., drift diffusion model, Ratcliff, 1978). It has been shown that when making such judgments, well-trained human participants can adaptively adjust this threshold to maximize reward rate (Balci et al., 2011; Simen et al., 2009) as well as accuracy (Bogacz, Hu, Holmes, & Cohen, 2010), and optimally integrate multiple pieces of information (Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014). However, prior studies concerned with comparing cue-based decision strategies under different time pressures have dealt exclusively with rather slow-paced decision scenarios, where time pressure conditions could range from 15–50 s (Bergert & Nosofsky, 2007; Payne et al., 1988; Payne, Bettman, & Luce, 1996; Rieskamp & Hoffrage, 2008). Other studies that required faster decisions (deadlines ranging between 600 ms to 1,600 ms) either extensively trained subjects with exemplars (Lamberts, 1995) or used choice problems that elicited participants' prior knowledge (Pachur & Hertwig, 2006), both of which encourages the use of memory-based heuristics. These circumstances, however, do not always approximate the kind of time pressure often faced in everyday and high-stakes decision-making, like in traffic, medical, or military scenarios, where choices have to be made by actively integrating multiple cues within a fraction of a second.

The goal of the present study, therefore, was to quantify the adaptive use of satisficing heuristics in an environment that approximates both the uncertainty and high-paced nature of much real-world decision making. To model real-life uncertainty, we eschewed the use of explicit cue values and instead developed a statistical learning task, inspired by the well-known “weather prediction task” (Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, & Gluck, 1994) that has been widely used to study learning and memory. In that task, participants are shown one, two, or three cards from a set of four “cue cards” (containing sets of abstract symbols), each of which predicts outcome of rain or sunshine with a given probability that is initially unknown to the participants. On each trial, participants are asked to predict the weather based on a

specific combination of cards and receive feedback, which enables them to gradually learn the probabilistic cue-outcome associations. However, due to small number of possible combinations of cue cards, the task is potentially susceptible to memorization of specific patterns that can discourage participants from actively combining evidence in each experimental trial (Yang & Shadlen, 2007). To overcome this limitation, instead of using four cues associated with fixed outcome probabilities, we presented two compound stimuli consisting of combinations of four binary cues on each trial (see Method section). Participants were asked to predict the stimulus that was likely to “win” as quantified by the combined weights of the cues that comprised the stimuli, which had to be learned via probabilistic feedback throughout the experiment. The large number of possible cue combinations (120 unique trials) prevented participants from memorizing specific patterns of stimuli-outcome combination, instead encouraging them to integrate available information to solve the task, allowing us to track and quantify participants’ use of cue information.

To gauge decision strategies under severe time constraints, we imposed decision time pressures in the subsecond range. Specifically, we first report two experiments, involving two independent cohorts of participants that each performed the same category learning task using compound (or integrated) cues: following an initial learning period, we assessed and compared decision making between two postlearning task phases, one unpressured phase (2-s response window), which was common to both experiments, and a subsequent high time pressure phase, where the degree of time pressure differed between the two groups of participants, ranging from moderate (Experiment 1: 750-ms response window) to severe (Experiment 2: 500-ms response window). We then employed variational Bayesian inference and Bayesian model selection analyses to infer the subjects’ decision strategies under the different time pressure conditions. Participants may feasibly approach this task in a number of different ways, including considering a random subset of cues, using a subset of cues with strong cue weights, using all the available cues, engaging in a take-the-best strategy, and so forth. To explore a large space of plausible decision strategies, we developed and contrasted 16 different plausible strategy models, allowing us to systematically track how participants integrate available information under changing time pressure. Then, to test generalizability of our findings, we conducted a third experiment using noncompound cues with a 3 s unpressured phase followed by a 750 ms time pressure phase. In sum, we used a large set of abstract cue stimuli combined with probabilistic feedback to infer how people learn to use multiple cues in the presence of uncertainty and how this usage changes as a function of high time pressure. The results establish that, under split-second time pressure, humans satisfice decision making by strategically discounting (or ignoring) the least informative cues.

Experiment 1: Satisficing Under Moderate Time Pressure (750 ms)

Method

For all experiments, we have reported all measures, conditions, data exclusions, and how we determined our sample sizes.

Participants. Forty-eight volunteers were recruited online through a human intelligence task (HIT) via Amazon Mechanical

Turk (MTurk). Assuming a medium-to-large effect size, we calculated that 38 usable data sets would provide us with a power of 0.8 at a Type I error level of 0.01 (Cohen, 1992). We recruited 10 extra participants in anticipation that about 10% to 20% of respondents would fail to meet performance criteria. All participants provided informed consent in line with Duke University institutional guidelines. To ensure high data quality, we followed MTurk’s recommended inclusion criteria of only inviting participants who had previously completed ≥ 500 HITs with a HIT approval rate of $\geq 90\%$ to participate in the experiments. We did not have any restrictions in age, demographics, or performance in approving the HIT, but assignments with more than 75 invalid trials ($>10\%$ of total number of trials) were rejected. Five participants were excluded from further analysis due to chance-level performance gauged by the percent optimal responses in the initial 480 trials prior to the time pressure phase (one-tailed binomial test, $p > .05$), leaving 43 participants (mean age = 34.4 years, $SD = 10.5$, 21–66 years; 17 female, 26 male). Participants were compensated with \$5.00 upon completion of the experiment, which lasted approximately 1 hr. In addition, a bonus payment of \$5.00 was given to the participant who earned the highest point score.

Stimuli. The task stimuli consisted of compound cues, constructed using four different features (color, shape, contour, and line orientation), which we refer to as *cue dimensions* (see Table 1). Each cue dimension was binary, comprising two subfeatures or *cue states*. For instance, the cue dimension of color had the two possible cue states of blue and red. Each cue state was associated with a fixed predictive value (or “weight”) indicating the probability of “winning”, and these values were complementary and summed to one within each cue dimension. For example, in a given participant, the color blue might have a weight of 0.6, which was its probability of winning, with the color red having a weight of 0.4. The “net weight” was the difference between the state weights for a dimension, in this case 0.2. It indicated how important that individual dimension was for selecting the winning stimulus (in this case, relatively unimportant). Table 2 displays the possible cue weights assigned across different cue dimensions. As can be inferred from the net weights, we created a compensatory environment, in which the highest cue, c_1 , can be out-weighted by some combinations of less valid cues. The weights were randomly assigned to the different cue dimensions for each participant at the beginning of the experiment but always followed the organization of Table 2. By exhaustively combining all possible cue states, 16 unique compound cue stimuli were constructed (Figure 1A). As explained in detail below, the sum of the weights associated with the four cue states comprising each stimulus governed the probability of that stimulus being a “winning” stimulus.

Table 1
Stimulus Organization

Cue dimension	Cue state	
	1	2
Color	Blue	Red
Shape	Circle	Square
Contour	White	Black
Line orientation	Horizontal	Vertical

Table 2
Assigned Cue Weights

Cue dimension	Cue state 1	Cue state 2	Net weight
c_i	$w_{c_{i,1}}$	$w_{c_{i,2}}$	$w_{c_{i,net}} = w_{c_{i,1}} - w_{c_{i,2}}$
c_1	.9	.1	.8
c_2	.8	.2	.6
c_3	.7	.3	.4
c_4	.6	.4	.2

Note. Net weights indicate the relative importance of each cue dimension in determining the winning stimulus.

Task. Participants performed a probabilistic classification task, in which they were required on each trial to compare two stimuli and select the one they deemed to be most likely to win (i.e., to have a higher total value than the other stimulus) by means of a time-restricted button press (Figure 1B). Stimuli (150×150 pixels each) were presented on the left and the right side of the screen (window size: $1,000 \times 700$ pixels) along the horizontal meridian, at an eccentricity of 250 pixels from a central fixation cross. The stimuli were sampled from the full set of the 16 compound cues (Figure 1A), such that each stimulus was paired with all the other stimuli except for itself, resulting in 120 unique trials. For a given trial, the compound cue stimuli could thus differ in one, two, three, or four cue dimensions. Stimuli were presented on the screen until a response was made or for the duration of an assigned response window of 2 s or 750 ms depending on experimental phase. Once a response was made, probabilistic feedback (see below), consisting of the words “win” or “lose,” was displayed for 1 s, followed by a 1-s intertrial interval (Figure 1B). A trial was considered invalid if a response was not made within a given response window or was made faster than 150 ms poststimulus. At the beginning of the experiment, participants were informed that they needed to learn about the values (weights) associated with the different cue states by trial-and-error, in order to collect as many points as possible. Participants earned 1 point for every winning trial and the total score at the end of the experiment was used to select the participant who received a bonus payment.

The probability that a left (L) or a right (R) stimulus would win was determined based on the cue states comprising the left (l) and the right (r) stimuli, $C = \{c_{1,l}, c_{2,l}, c_{3,l}, c_{4,l}, c_{1,r}, c_{2,r}, c_{3,r}, c_{4,r}\}$, and their associated weights, $\{w_{c_{1,l}}, w_{c_{2,l}}, w_{c_{3,l}}, w_{c_{4,l}}, w_{c_{1,r}}, w_{c_{2,r}}, w_{c_{3,r}}, w_{c_{4,r}}\}$ (see Table 2). The difference in the cue weights along each dimension governed the winning probability, as described in the equations below (adopted and modified from Yang & Shadlen, 2007):

$$P(L|C) = \frac{10 \sum_{i=1}^4 (w_{c_{i,l}} - w_{c_{i,r}})}{1 + 10 \sum_{i=1}^4 (w_{c_{i,l}} - w_{c_{i,r}})} \quad (1)$$

$$P(R|C) = 1 - P(L|C)$$

where i represents cue dimension. For example, if there was no difference in overall weights of each stimulus, $\sum_{i=1}^4 (w_{c_{i,l}} - w_{c_{i,r}}) = 0$, then $P(L|C) = P(R|C) = .5$. At the other extreme, if the left stimulus consisted of cues having every one of the higher cue states (see Table 2), the probability that the left stimulus would win could be calculated as, $P(L|C) = 10^2 / (1 + 10^2) = 0.99$. Based on Equation 1, feedback was determined probabilistically

on a trial-by-trial basis. Thus, because feedback was probabilistic, it was not providing participants with “correct” feedback in an absolute sense, because there could be situations where a participant might receive negative feedback for an objectively correct decision.

Our goal was to first allow participants to learn the (uncertain) cue values, and to then compare decision making strategies between conditions of low versus moderate time pressure. To this end, participants completed three different phases of the probabilistic classification task: (a) an initial learning phase, followed by (b) a no time pressure (NP) phase, followed by (c) a time pressure (TP) phase to create satisfying pressure. In each phase, participants completed 240 trials, consisting of the full set of possible trials presented twice, where stimuli in the second set were presented in the opposite locations (i.e., as mirror images) to the first set. Trials were grouped into 12 blocks of 60 trials each, with short breaks in between. In both the initial learning period and NP phases, participants were given a maximum of 2 s to respond. Then, following the NP phase, participants performed an additional 240 trials of the TP phase, in which participants were given a 750-ms response window (moderate time pressure). This time pressure was based on pilot work with this task, where, without time pressure, we obtained mean response times of around 700 ms. Hence, Trial Sets 1 and 2 made up the learning phase, Sets 3 and 4 comprised the NP phase, and Sets 5 and 6 formed the TP phase. Probabilistic feedback was provided throughout all three phases of the experiment. Our analyses focused on assessing and contrasting decision making strategies in the experimental (postlearning) phases, that is, in the NP and TP phases.

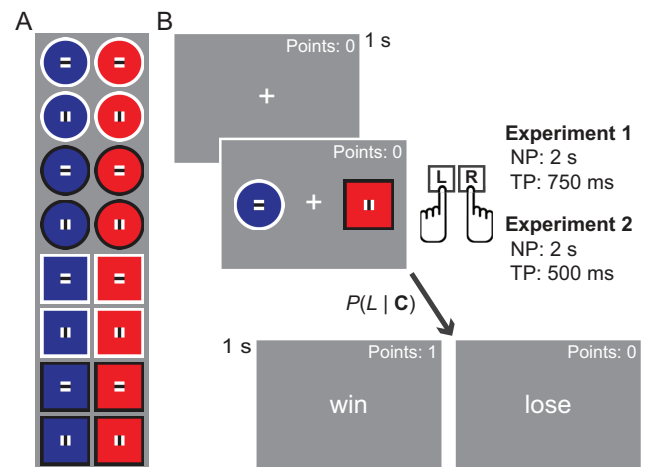


Figure 1. Stimuli and task of Experiments 1 and 2. (A) We constructed 16 unique compound stimuli by combining 4 cue dimensions (color, shape, contour, line orientation), each with binary cue states (e.g., red vs. blue) of varying “weights” (see Tables 1 and 2). (B) On each trial, participants performed a probabilistic classification task by selecting the stimulus they deemed more likely to “win,” with varying response windows of 2 s (initial learning, and NP phase), 750 ms (TP phase in Experiment 1), and 500 ms (TP phase in Experiment 2). Feedback was determined probabilistically based on the probability that a left (L) stimulus would win given cue states (C) presented on each trial, $P(L|C)$ (see Eqn. 1), and displayed for 1 s upon each response. NP = no time pressure; TP = time pressure. See the online article for the color version of this figure.

Survey. To gauge participants' knowledge of cue values and decision strategies, at the end of the experiment, participants filled out a survey, which included multiple-choice questions about participants' beliefs concerning the cue rankings, number of cues used in each phase, and cue-outcome probabilities. First, participants were asked to rank cue dimensions in order of their perceived importance in predicting outcomes: "Please rank the cues from the most informative (1) to the least (4). For example, if you thought color cues (blue/red) were the most reliable indicator for winning a given trial, or if you've made decisions mostly based on color cues, color would be the most informative cue." Then, they were asked to indicate how many cue dimensions they considered in making their decisions during the NP and the TP phase: "In the slow phase, how many cues, on average, did you take into account before making decisions?" and "In the fast phase, how many cues, on average, did you take into account before making decisions?" "Self-report strategy models" (as opposed to objectively inferred ones) were constructed based on the answers to these questions. For example, if a participant ranked cues as color, shape, contour, and line orientation (ordered from the most to least informative) and claimed to have used two cues during the TP phase, then his or her self-report TP model was a model that included only color and shape cues. In addition, to check whether participants learned the relative cue state values correctly within each cue dimension, we asked them for each cue dimension to select the cue state with the higher value and to estimate that value (winning probability), along a range from 50% to 100%. For instance, for the color dimension, they were asked "Which cue [state] has a higher probability of winning?" and provided with a choice between blue and red with pictures of two stimuli differing only in the color dimension. Participants were then asked "What is your estimate of a winning percentage of the cue [state] you've selected above?" which they had to indicate on a scale ranging from 50% to 100%.

Data analysis. Data analyses were based on optimal choices that were favored by the cue weights, independent of the actual (probabilistic) feedback. In other words, whenever a participant chose the stimulus with the higher probability of winning according to Equation 1, the trial was considered "correct" (optimal) even though the probabilistic nature of the feedback could have resulted in negative feedback for that particular trial. For the purpose of evaluating percent optimal choices, trials with two stimuli that had an equal sum of weights were excluded, because a correct choice cannot be defined in these trials. Hence, if a participant learned the cue weights optimally, as in the case of an ideal observer (see below), and correctly integrated them, that participant could in theory achieve 100% accuracy (which would not be associated with 100% "win" feedback though). However, it has been shown in a number of previous studies that when receiving probabilistic feedback, people tend to match their choice probabilities to the outcome probabilities (see, e.g., Vulkan, 2000). From this matching perspective, performance with 79% accuracy represents the ideal percentage of optimal choices in our protocol. In reporting analysis of variance (ANOVA) measures, violations of sphericity assumptions were corrected by Greenhouse-Geisser correction to the degrees of freedom. Similarly, in reporting *t* test results, degrees of freedom were corrected for unequal variance where necessary.

Ideal observer model. To identify the optimal performance level for the different task phases, we employed an ideal observer

model. An ideal observer was exposed to the same trial sequences and feedback as participants and learned the cue weights optimally throughout the NP phase. This approach enabled us to estimate the "ideal" cue weights that participants could have learned based on the probabilistic feedback that they received.

Subjective cue weight analysis. As noted above, the net weights (see Table 2) correspond to the objective importance of each cue dimension in identifying the "winning" stimulus. We assumed that a decision variable (DV) is computed by cue weights that a given subject learned over time such that the subject chooses left, when $DV > 0$, and right, when $DV < 0$. Ideally, if the subject had learned all the cue weights correctly, as in the case of the ideal observer model, the corresponding psychometric function would be a step function. However, people tend to make mistakes, which can be better modeled using an S-shaped function that accounts for decision errors (see Figure 3A and B). This is equivalent to adding a little bit of noise into the DV or sampling the weights from a distribution of subjective beliefs. Hence, to assess the post hoc, subjective importance of each cue dimension for each subject, we performed a logistic regression analysis. We estimated the effect of individual cue dimension on subjects' choices throughout the total number of trials, N , in each phase, as follows:

$$P_{Left} = \frac{1}{1 + e^{-(B_0 + XB)}} \quad (2)$$

$$\text{where } X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,4} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,4} \end{bmatrix}, B = [\beta_1 \beta_2 \beta_3 \beta_4]^T$$

with B_0 representing a $N \times 1$ matrix for estimating the intercept. Each element of matrix X on i th cue dimension of n th trial was defined as following:

$$x_{n,i} = \begin{cases} -1 & \text{if } w_{i,l} < w_{i,r} \\ 0 & \text{if } w_{i,l} = w_{i,r} \\ 1 & \text{if } w_{i,l} > w_{i,r} \end{cases} \quad (3)$$

Subjective cue weights, w_i^* , were calculated by transforming the four fitted coefficients, β_i , to log base 10:

$$w_i^* = \log_{10} e^{\beta_i} \quad (4)$$

The magnitude of w_i^* also roughly corresponds to the decision noise so that a subject may be said to be relatively optimal when their subjective weights are proportional to true weights.

Similar to the generalized take-the-best model (gTTB; Bergert & Nosofsky, 2007), this approach is a probabilistic generalization of the cue usage based on log odds instead of a linear combination of cue weights. Although both models yield roughly similar predictions for the probability of choosing the left stimulus, P_{Left} , the main discrepancy between the two models is that when all cues presented on a given trial (ranging from 1 to 4 cues) are in favor of the left (or right) stimulus, the gTTB will predict $P_{Left} = 1$ (or $P_{Left} = 0$) whereas the logistic model will predict the probability scaled by the sum of cue weights. For consistency with our probabilistic feedback protocol, which was also based on the sum of cue weights (see Equation 1), we chose to adopt a logistic regression analysis instead of the gTTB to achieve better sensitivity in inferring subjective cue weights. Another popular model of decision noise is a "lapse model" that assumes random guessing in some of the trials. However, this model is not appropriate for

characterizing the participants' behavior, because a lapse model predicts psychometric curves that do not saturate at 0% or 100% when the sum of evidence is at the minimum or maximum value. This is not consistent with our empirical observations (again, see Figure 3A and B). We therefore did not attempt to fit lapse model variants to our data. We also note that, while we have not explicitly considered a strategy which is a mixture of strategies considered, the Bayes factors that we do compute (see below) can be used to approximately determine the mixture proportions on a per subject basis (see Friston et al., 2016 for details).

Decision strategy models. We explored a large space of plausible decision models that participants may have applied to solve the task, resulting in the comparison of 16 different models (Figure 4A) to identify participants' decision strategies under conditions with and without time pressure. Models 1 to 15 were created by accounting for every possible case of cue usage, with Model 15 serving as an optimal cue-integration model. In other words, Model 15 represented a compensatory strategy model where participants were assumed to integrate the weights over all four cue dimensions in making their choice, whereas Models 1–14 consisted of all possible cue combinations short of the full integration over the four cue dimensions. In addition, Model 16 was constructed based on the take-the-best satisficing algorithm (Gigerenzer & Goldstein, 1996), which searches through the cues in order of descending value until it finds the first cue that differentiates between the two stimuli. For example, although both Model 1 and 16 each use a single cue to arrive at a decision, Model 1 assumes that the subject uses only the highest value (most discriminatory) cue dimension, c_1 , and makes a random guess if this cue does not differ between the two stimuli, whereas Model 16 searches through the cue dimensions from c_1 to c_4 in descending value order until it finds the highest-value cue that discriminates between the two stimuli. In Figure 4B, we plot the expected accuracy for each strategy model under the assumption that a decision maker has learned to make optimal choices. As expected, the optimal cue-integration model, Model 15, achieves the best performance, which further validates our compensatory cue structure.

Based on the strategy models, we conducted Bayesian model comparison, using variational Bayesian inference (Drugowitsch, 2013), which returns parameters of a fitted logit model, w_m with input matrix, x_m , and output choice vector, y , for a given strategy model, m :

$$P(y = 1 | x_m, w_m) = \frac{1}{1 + e^{(-w_m \times x_m)}} \quad (5)$$

$$P(w_m | \alpha_m) = \mathcal{N}(w_m | 0, \alpha_m^{-1} I) \quad (6)$$

$$P(\alpha_m) = \text{Gamma}(\alpha_m | a_0, b_0) \quad (7)$$

Hyper-parameters for the prior over weights ($a_0 = .01$ and $b_0 = .0001$) were chosen so that they corresponded to a very weak prior. This approach allows estimation of an approximate posterior distribution over the weights, and it marginalizes out the uncertainty to obtain a marginalized likelihood of the model, m . Hence, Bayes factors, BF_m , were computed for each model in comparison to Model 15 (the optimal, "full" cue integration model):

$$BF_m = \frac{P(D | Model_m)}{P(D | Model_{15})} \quad (8)$$

where D denotes observed data and m denotes model number. Any model that had BF_m greater than 3 was considered as having greater evidence in its favor than the optimal cue-integration model (Kass & Raftery, 1995). Using this approach, we were able to infer the most likely strategy that each participant may have employed to solve the task. Hence, the advantage of using such analysis is that even if participants did not learn optimally, it is possible to deduce the most likely cue usage based on each individual's choice patterns.

In a control analysis, we established that changing the hyper priors based on the posterior distributions of the weights computed from the initial learning phase did not yield any significant influence on Bayes factors nor model comparison results (see Figure 4). More precisely, consistent with an empirical Bayesian approach, the hyper priors, a_0 and b_0 , were estimated based on the mean (μ) and the variance (σ^2) of the observed distribution of subjective cue weights in the initial learning phase:

$$a_0 = \frac{\mu^2}{\sigma^2}, \quad b_0 = \frac{\mu}{\sigma^2} \quad (9)$$

which yielded the following values: $a_0 = .594$ and $b_0 = .815$ for Experiment 1, $a_0 = .501$ and $b_0 = .798$ for Experiment 2, and $a_0 = .479$ and $b_0 = 1.804$ for Experiment 3.

Bayesian model selection. To characterize strategy usage at the group level, we used the Bayesian model selection approach of Stephan, Penny, Daunizeau, Moran, and Friston (2009), which treats models as random effects that could vary across subjects and estimates the parameters of a Dirichlet distribution to obtain the probabilities for all strategy models considered. These probabilities are then used to define a multinomial distribution, which can be used to estimate the probability that model m generated the data of each subject, as well as the exceedance probability. The exceedance probabilities reported here were calculated by summing the approximate log model evidences to the spm_BMS routine of the SPM8 software suite (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>).

Results

Task performance. Prior to analyzing how subjects weighted the cues and arrived at their decisions, we briefly summarize general task performance. Mean proportion of optimal choices and response time (RT) for each set and phase are shown in Table 3. Here, an optimal choice refers to the "correct" decision favored by the sum of cue weights, independent of the actual feedback provided. As can be seen in Figure 2 (black line), participants gradually learned to choose the higher-value stimuli in the learning phase, achieving 72.1% optimal responding by Set 2. We observed no significant difference in performance between Set 2 and 3 ($t(42) = 0.1, p = .92$), suggesting that performance had stabilized by the end of the learning period. However, percent optimal choices increased from Set 3 to Set 4 ($t(42) = 3.4, p = .001$), thus suggesting that some residual cue learning was still taking place during the NP phase. This residual learning effect is not surprising since probabilistic feedback was provided throughout the experiment. As expected, the percentage of optimal choices and RT

Table 3
Summary of Task Performance in Average Percent Optimal Choices and RT in ms

Phase	Set	Experiment 1		Experiment 2		Experiment 3	
		% optimal choices	RT (ms)	% optimal choices	RT (ms)	% optimal choices	RT (ms)
LP	1	66.5 (1.6)	732 (27)	63.6 (1.8)	749 (37)	62.9 (1.5)	1150 (61)
	2	72.1 (1.5)	702 (24)	70.4 (1.4)	729 (33)	69.4 (1.6)	1122 (53)
NP	3	72.0 (1.4)	682 (23)	68.9 (1.8)	701 (34)	72.8 (1.3)	1126 (52)
	4	75.4 (1.5)	658 (21)	73.6 (1.7)	674 (32)	74.5 (1.5)	1076 (52)
TP	5	71.4 (1.4)	457 (8)	62.6 (1.5)	359 (7)	62.7 (1.4)	510 (13)
	6	73.2 (1.8)	458 (8)	65.1 (1.9)	367 (6)	63.2 (1.6)	511 (11)

Note. Numbers in parenthesis are standard error of the mean (SEM). Experiment 1 represents moderate pressure group with 750-ms response window in the TP phase; Experiment 2 represents high pressure group with 500-ms response window in the TP phase. Both Experiments 1 and 2 used compound stimuli with 2-s response window in the initial learning (LP) and NP phases; Experiment 3 represents noncompound stimuli group with 3-s response window in the LP and NP phases and 750 ms window in the TP phase. RT = response time; TP = time pressure; NP = no time pressure.

reliably scaled with the difficulty of the decisions, as defined by the difference in sum of cue weights between the two stimuli on each trial. More difficult decisions (smaller weight differences) were associated with decreasing percent correct choices (linear trend: $F(1, 42) = 322.1, p < .001$) and increasing RT (linear trend: $F(1, 42) = 45.2, p < .001$).

In the subsequent TP phase (Figure 2, black line), participants experienced a moderate time pressure of 750 ms, which was close to (but greater than) the mean RT in the NP phase (see Table 3). Here, the rate of optimal choices did not differ significantly from the NP phase ($t(42) = 1.6, p = .11$), but responses were accelerated ($t(42) = 11.5, p < .001$), indicating that participants modified their approach to the task to adapt to the higher time pressure. We found no significant difference in performance between Set 5 and 6 ($t(42) = 1.5, p = .15$). Similar to the NP phase, both percent correct choices (linear trend: $F(1, 42) = 137.8, p < .001$) and RT (linear trend: $F(1, 40) = 117.1, p < .001$) were modulated by

decision difficulty. Finally, assessing the potential impact of basic individual differences on task performance, we found no effect of gender on choices in either task phase (NP: $t(41) = .9, p = .38$; TP: $t(41) = 1.1, p = .29$), and no correlation between age and performance (Pearson's correlation, NP: $r = -.2, p = .19$; TP: $r = -.1, p = .51$). In sum, these initial analyses demonstrate basic statistical learning of our task as well as an effect of time pressure on RT. To examine potential shifts in decision making strategies under time pressure, we next turned to determining participants' cue weighting and cue-integration strategies in performing the task.

Subjective cue weights. To assess the extent to which each cue dimension (e.g., color) affected participants' decisions, we performed logistic regression (Equations 2–4, Figure 3) based on their performance in the NP and TP phases. The output of the logistic regression analysis corresponds to log odds of choosing the left stimulus given the presence of differing cue states, and thus, provides a way to measure how much subjective net weight was assigned to each of the four cue dimensions (as contrasted with the objective, a priori established net weights; see Table 2). Logistic regression curves for an example participant are displayed in Figure 3A and B.

At the population level (Figure 3C), an ANOVA with assigned cue weights (w_i) and experimental phase revealed a significant main effect of cue weights ($F(2.2, 91.98) = 20.3, p < .001$), which was characterized by a significant linear trend ($F(1, 42) = 33.7, p < .001$), suggesting that participants had correctly learned the relative ranks of the cues. There was no effect of NP/TP phase ($F(1, 42) = 2.0, p = .17$) and no Cue Weight \times Phase interaction ($F(3, 126) = .02, p = .99$). Subjective cue weights of each participant were further compared to weights produced by an ideal observer, by computing Spearman's rank correlation coefficients. An ideal observer was exposed to the same stimulus and feedback sequences and hence, ideal weights reflect weights that participants would have acquired if they had learned optimally. Table 4 shows a summary measure of correlation coefficients averaged across participants (tested for significance using standard t tests). All the mean correlation coefficients were significantly greater than zero, confirming positive correlations between participants' subjective cue weights and the ideal weights. In sum, the subjective cue weight analysis indicates that participants learned to accurately rank the informational values of the four cues. Next, we

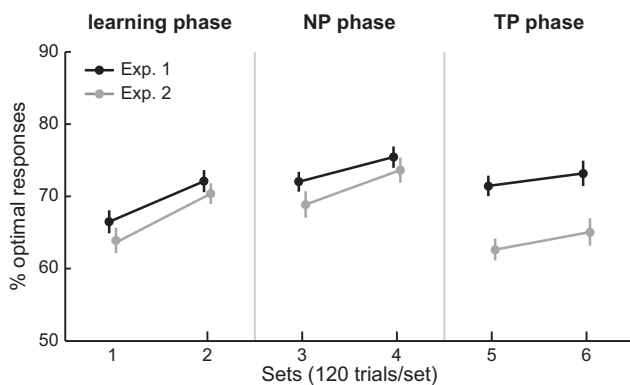


Figure 2. General task performance of Experiments 1 and 2. Average percent optimal responses in Experiment 1 (750 ms moderate time pressure, black line) and Experiment 2 (500 ms severe time pressure, gray line) in the learning (left), NP (middle) and TP (right) phases. An optimal choice refers to the "correct" decision favored by the sum of cue weights, that is, choosing the stimulus with the higher probability of winning, independent of the actual feedback provided. Trials with two stimuli with equal sum of weights were excluded. Error bars are standard errors (SEM). NP = no time pressure; TP = time pressure; Exp = Experiment.

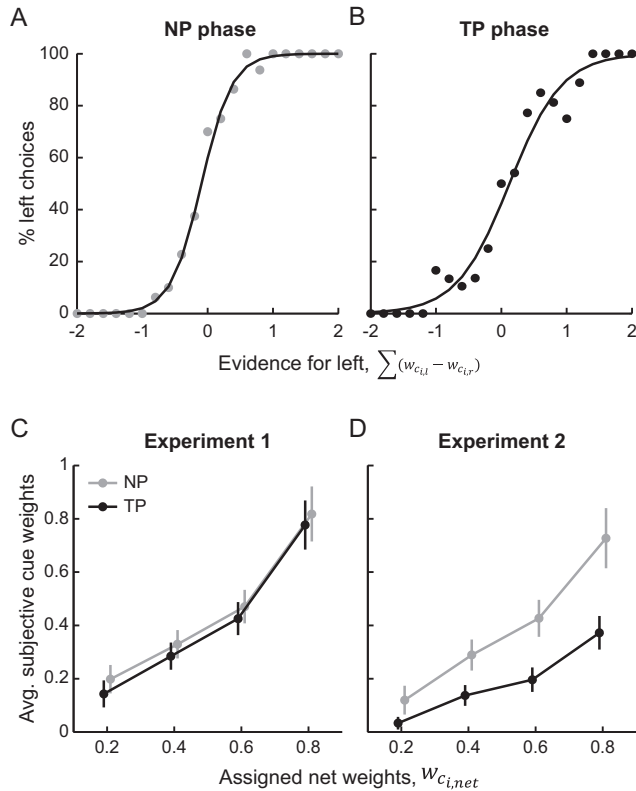


Figure 3. Logistic regression fit of individual performance and estimation of subjective cue weights (A, B). Performance of a single example participant in the (A) NP (2-s response window) and (B) TP (750-ms response window) phases. The fraction of left choices is plotted as a function of amount of evidence in favor of left stimuli, represented by differences in sum of cue weights. An individual data point reflects percentage of left choices made given the sum of evidence (y-axis). Curves are logistic regression fits to the data (Equations 2–4). (C, D) Average subjective cue weights over groups of subjects as a function of NP (gray line) and TP (black line) phases, for (C) Experiment 1 (750 ms TP) and (D) Experiment 2 (500 ms TP). Error bars are standard errors (SEM). NP = no time pressure; TP = time pressure.

determined how participants combined the cue weights in reaching their decisions.

Decision strategy model comparison and selection. To quantify participants' decision-making strategies, we used variational Bayesian inference (Drugowitsch, 2013) to gauge evidence for different decision models in reference to the optimal cue-integration model, Model 15 (Figure 4A; see Method section). A marginalized likelihood of each model per participant was then used to characterize overall decision strategies at the group level through Bayesian model comparison (Stephan et al., 2009). In the NP phase, the optimal cue-integration model (Model 15) was the winning model with exceedance probabilities of .78 (Figure 4C). Under 750 ms time pressure (TP phase), however, a model that used only the three most predictive cues (Model 11, using c_1 , c_2 and c_3) was the most likely model with an exceedance probability of .67 (Figure 4C). The additional control analysis using informative priors estimated based on the posterior distributions of the weights from the initial learning phase yielded highly similar

results (Figure 4E). These results suggest that participants shifted from using all four cues when having no time pressure to using only the three highest-value cues under moderate time pressure of 750 ms.

Taken together, the subjective cue weight and decision model analyses suggest that (a) participants learned to correctly rank the values of the cue dimensions, and (b) under moderate time pressure, they disregarded the least valuable cue dimension. We reasoned that the latter “dropping” of the worst cue dimension from the decision process under moderate time pressure could reflect one of two processes: it could either reflect a strategic shift in processing, whereby participants categorically ignore the worst cue dimension in their decision making, or it could simply be due to the fact that participants are running out of time in a serial, value-ranked cue integration process. In the latter scenario, participants would still attempt to use the worst cue, but on most trials they do not have sufficient time to process it, because they first attend to higher valued cues. The fact that the average RT for these trials in the TP phase was well below the response deadline of 750 ms ($mean = 481$ ms; one-sample t test: $t(42) = 29.2$, $p < .001$) speaks against the hypothesis that participants simply ran out of time in considering this cue.

However, to adjudicate between these two possibilities in greater detail, we analyzed choices in trials where the worst cue was the *only* distinguishing dimension between the two stimuli (16 trials/phase). In these trials, participants should be able to evaluate that cue dimension, since they do not have to spend time on integrating differential values over the other cue dimensions, and the single differentiating visual feature should be quite salient perceptually. Thus, if subjects dropped the worst cue in the TP phase due to a lack of time in serial cue integration, they should nevertheless perform above chance on these trials. By contrast, if subjects strategically disregarded the worst cue dimension under time pressure, they would simply guess on these trials. This analysis revealed that 30 out of 43 participants (70%) performed no different from chance level (binomial test, $p > .05$) when only the least important cue differentiated the stimuli, indicating that majority of participants guessed in these trials. To further corroborate this result, we computed Bayes Factors (BF) comparing two models, with H_0 : the probability of correct choice, $p = .5$ and H_1 : all possible values of p in $[0,1]$ is equally probable. This analysis revealed that 28 out of 43 participants (65%) had a $BF > 1$ favoring H_0 although only seven participants showed a significant effect ($BF > 3$; Kass & Raftery, 1995). At the group level, we estimated BFs for use in a paired t test (adopted from Rouder, Speckman, Sun, Morey, & Iverson, 2009) based on average performance between the NP ($mean = 58.9\%$, $SEM = 3.5$) and TP ($mean = 59.6\%$, $SEM = 3.1$) phases. This t test did not show any significant difference ($BF = 5.88$, favoring the null hypothesis). In sum, we obtained equivocal results in this selective analysis of low-value cue decisions, which do not allow us to draw strong conclusions about whether participants truly performed at chance on these trials. Note though that power in these analyses is limited, as we only considered a small number of trials (16 trials per phase) per participant.

Comparison of the TP phase results between the estimated subjective cue weights (Figure 3C) and the strategy model selection (Figure 4C and E), at a glance, may seem odd because the

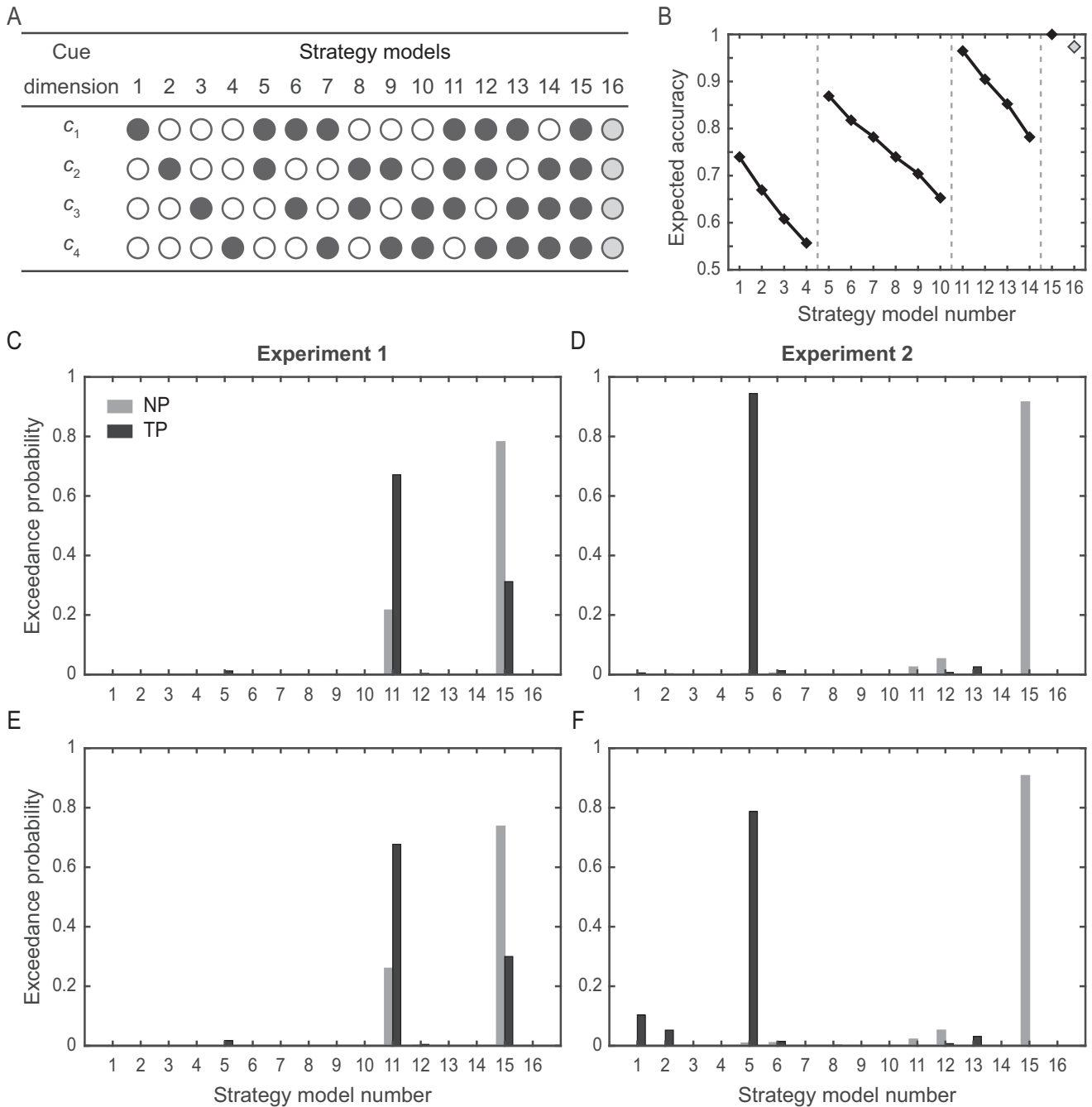


Figure 4. Decision strategy models and Bayesian model selection. (A) Strategy models were constructed by accounting for every possible case of cue usage (Model 1–15) with Model 15 representing an optimal cue-integration model. Filled circles (●) denote the cues that are included in a given model. Model 16 was constructed based on the take-the-best heuristic (Gigerenzer & Goldstein, 1996). (B) Expected accuracy, determined by the percentage of correct choices independent of the probabilistic feedback, was estimated as a function of cue usage defined by strategy models. The optimal model, Model 15, achieves 100% accuracy given that a decision maker has learned the cue weights optimally. Dotted lines separate strategy models according to the number of cues used (C–D). Bayesian model selection group results for the NP (gray) and TP (black) phases of (C) Experiment 1 (750 ms TP) and (D) Experiment 2 (500 ms TP) using uninformative priors. Exceedance probability represents how likely a particular model is given the group data (Stephan et al., 2009). (E–F) To control for the effect of priors, Bayesian model selection analysis was repeated using informative priors based on the posterior distributions of the weights computed from the initial learning phase of (E) Experiment 1 and (F) Experiment 2. NP = no time pressure; TP = time pressure.

Table 4
Mean Rank Correlation Coefficients (Spearman's r_s) Between Ideal Weights, Subjective Weights, and Explicit Survey Ratings for Experiment 1, 2, and 3

Experiment	NP	TP
Experiment 1		
Ideal - Subjective	.48**	.45**
Ideal - Survey	.45**	.44**
Subjective - Survey	.54**	.48**
Experiment 2		
Ideal - Subjective	.55**	.45**
Ideal - Survey	.31*	.27*
Subjective - Survey	.37**	.41**
Experiment 3		
Ideal - Subjective	.55**	.50**
Ideal - Survey	.61**	.54**
Subjective - Survey	.67**	.71**

Note. Ideal = ideal weights produced by an ideal observer; Subjective = subjective weights computed based on participants' behavioral data (see Figure 3); Survey = participants' self-reported cue ranks. NP = no time pressure; TP = time pressure.

* $p < .05$. ** $p < .001$ from one-sample t -test.

population average of the weight of the least important cue (c_4) does not drop to zero, while the modeling results suggest that these cues were dropped from the decision making process. It should be noted though that although the majority of participants learned the correct cue rankings, there was a lot of variability in estimated cue weights across subjects. Given that the magnitude of cue weights also roughly corresponds to the amount of decision noise in each participant, fitted subjective cue weights were also highly dependent on individual task performance. The combination of an increase in decision noise and a decrease in number of observations in the TP phase, therefore, may have contributed to the increased variability in the subjective cue weights. To rule out the possibility that this increase in noise affecting the estimation subjective cue weights under time pressure is what is driving our group results of the Bayesian model selection, we reran the analysis using only the least noisy half of our data set, that is "high performing" subjects, based on a median split of performance accuracy in the initial learning phase (average performance $\geq 65\%$). The results of this analysis remained the same as for the whole sample (data not shown). Furthermore, we observed a significant down-weighting of the least important cue in the TP phase compared to the NP phase ($t(20) = 2.5, p < .05$), whereas no significant difference was observed for the rest of the cue weights.

Overall, our decision strategy analysis provides evidence that participants engaged in optimal, exhaustive cue value integration when time pressure was low. When faced with moderate time pressure, however, our analyses provided some evidence indicating that the participants satisficed strategically through cue discounting, compensating for the lack of time via a "drop-the-worst" cue strategy, that is, they ignored the least predictive subset of the cue information. As shown in Figure 4B, switching from the optimal cue integration model, Model 15, to Model 11 leads to only a negligible difference in expected accuracy (97%), which further supports the adaptive nature of strategy selection under moderate time pressure.

Survey results. To gauge the degree to which decision-making was driven by explicit knowledge concerning the different cue states, we analyzed a range of survey questions. When participants were asked to define a cue state with higher weight within a cue dimension (e.g., blue vs. red), the mean number of correct responses (out of a possible maximum of 4) was 3.4, indicating that the majority of participants were able to explicitly understand the relative importance of the cue states within each dimension. Then, to identify a "self-report strategy model" for each participant (see Method section), we surveyed relative ranks of cue dimensions and number of cue dimensions considered in each phase. As can be seen in Table 4, participants' ratings of relative rank of cue dimensions, from the most to the least informative, showed a significant positive correlation with (a) the subjective cue weights estimated using logistic regression analysis, and (b) the ideal weights estimated from the ideal observer model.

In terms of number of cue dimensions used in each phase, participants answered that they used significantly fewer cues ($t(42) = 7.8, p < .001$) during the TP phase ($mean = 1.8, SD = .7$) compared with the NP phase ($mean = 2.6, SD = .7$). These self-report models, however, did not match well with our findings based on the objectively inferred strategy models. That is, only five participants out of 43 were able to correctly identify their own decision strategies for both the NP and TP phases. In sum, the survey results indicate that participants gained some explicit knowledge of the relative cue weights; however, they did not have much insight into the strategies they employed in translating these cue weights into decisions.

Discussion

In order to characterize satisficing decisions under uncertainty and split-second time pressure, we developed a new multicue statistical learning protocol and applied a range of analyses to infer participants' learning of cue values and their use in the decision process. The results of Experiment 1 document that participants reliably acquire knowledge of cue values from trial and error learning, and Bayesian model selection suggests that they employ a decision strategy of (optimal) exhaustive cue information integration when they are not under time pressure (2-s response window); however, when put under moderate pressure (750-ms response window), participants appear to strategically adapt their decision process by ignoring the least valuable cue dimension. Thus, Experiment 1 seems to have revealed a novel satisficing strategy of "drop-the-worst" cue under time pressure. However, a confirmatory analysis focusing on those trials where only the lowest-value cue differentiated between the two stimuli produced equivocal results. The latter might be attributable to a small trial count ($n = 16$ trials/phase) and/or a subset of participants who were able to perform optimally even under moderate time pressure (see Figure 4C). Therefore, we followed up these initial findings with a second experiment that pursued two main goals. First, we aimed to replicate the decision making pattern observed in Experiment 1. Second, we sought to characterize decision making when time pressure was increased even further. If "drop-the-worst" is a reliable general strategy in split-second satisficing, then we would expect participants to further trim their usage of lower value cues as the time pressure increases. To test this hypothesis, we ran an exact replication of Experiment 1 in a new cohort of subjects, with

the only difference being that the TP phase in Experiment 2 was reduced, from 750 ms to 500 ms.

Experiment 2: Satisficing Under Severe Time Pressure (500 ms)

Method

Participants. A new, nonoverlapping cohort of 40 volunteers was recruited through a separate HIT on Amazon MTurk with the expectation of collecting 38 usable data sets as in Experiment 1. The same HIT approval criteria used in Experiment 1 were applied to ensure high data quality. Four participants were further excluded due to chance-level performance (one-tailed binomial test on percent optimal responses in the initial 480 trials without time pressure; $p > .05$), leaving a total of 36 participants (mean age = 32.0 years, $SD = 10.4$, 21 – 64 years; 15 female, 20 male, 1 unknown). Participants were paid \$5.00 upon completion of the experiment and an additional \$5.00 bonus payment was given to the participant who achieved the highest score. The age ($t(76) = 1.0$, $p = .31$) and gender ($\chi(1) = .1$, $p = .77$) distribution of participants in Experiment 2 was equivalent to those of Experiment 1.

Task. Experiment 2 was identical to Experiment 1, except that the response time window in the TP phase was reduced from 750 ms to 500 ms (less than the mean RT of the NP phase in Experiment 1) to observe how decision strategies change under more severe time pressure.

Results

Task performance. Similar to Experiment 1, participants were able to gradually improve their performance throughout the first 240 trials in the learning phase. We observed no significant difference in performance between Set 2 and 3 ($t(35) = 1.0$, $p = .31$), suggesting that performance had stabilized by the end of the learning period. However, as in Experiment 1, optimal choices increased from Set 3 to Set 4 ($t(35) = 3.8$, $p = .001$), suggesting that some residual cue learning was still taking place during the NP phase as participants continued to observe probabilistic feedback provided upon each choice (see Figure 2, gray line, and Table 3). As expected, in the NP phase, the percentage of optimal choices and RT were affected by decision difficulty with more difficult decisions resulting in a significant decrease in performance (linear trend: $F(1, 35) = 267.4$, $p < .001$) and increase in RT (linear trend: $F(1, 35) = 29.3$, $p < .001$). Also as expected, given the equivalence between the learning period and NP phases between Experiments 1 and 2, no difference in performance of Set 1 through 4 between experiments was observed (main effect of experiment: $F(1, 77) = 1.5$, $p = .23$; Experimental Group \times Set interaction: $F(3, 231) = .3$, $p = .81$), indicating comparable cue learning and NP performance across the two cohorts.

In the TP phase (Figure 2, gray line), however, both optimal responses ($t(35) = 5.5$, $p < .001$) and RT ($t(35) = 11.0$, $p < .001$) decreased significantly compared to the NP phase. Accordingly, the percentage of optimal choices in the TP phase was significantly lower in the severe pressure group than in the moderate pressure group (between-subjects effect of experiment: $F(1, 77) = 15.3$, $p < .001$). There was no difference in performance between Set 5 and 6 ($t(35) = 1.7$, $p = .10$). Despite a significant change in

performance, the effect of decision difficulty on percent optimal choices (linear trend: $F(1, 35) = 32.1$, $p < .001$) and RT (linear trend: $F(1, 29) = 68.6$, $p < .001$) was still present in this phase. Similar to Experiment 1, there was no significant effect of gender on task performance (NP: $t(33) = .2$, $p = .83$; TP: $t(33) = .7$, $p = .48$) as well as no significant correlation between age and performance (Pearson's correlation, NP: $r = .02$, $p = .90$; TP: $r = -.1$, $p = .42$). In sum, throughout the learning period and NP phases, which were equivalent between Experiment 1 and 2, participants successfully acquired the (uncertain) values of the different cue dimension via trial-and-error learning. However, there was clear evidence that the 500 ms time pressure in the TP phase of Experiment 2 had a detrimental effect on performance, both compared to the NP phase of Experiment 2, as well as to the 750ms TP phase in Experiment 1.

Subjective cue weights. Separate sets of subjective cue weights for the NP and TP phases were obtained using logistic regression (Figure 3D). An ANOVA revealed a significant effect of cue weights ($F(2.2, 76.9) = 14.2$, $p < .001$), with a significant linear trend of cue weights ($F(1, 35) = 27.7$, $p < .001$) indicating that participants learned the correct relative ranks of cue weights. The main effect of NP/TP phase ($F(1, 35) = 47.2$, $p < .001$), and the Cue Weight \times Phase interaction ($F(2.4, 84.1) = 4.1$, $p = .01$) were also significant. Post hoc analyses revealed significant main effects for both NP ($F(2.2, 76.2) = 12.3$, $p < .001$) and TP ($F(2.5, 86.0) = 11.0$, $p < .001$) phases on cue weights. In addition, there was a significant effect of time pressure phase on subjective cue weights ($F(1, 143) = 47.9$, $p < .001$), which was further confirmed by significant pairwise cue weight differences for cues 0.4 ($t(35) = 3.5$, $p = .001$), 0.6 ($t(35) = 4.0$, $p < .001$), and 0.8 ($t(35) = 4.5$, $p < .001$), demonstrating an overall down-weighting of cue values under severe time pressure. The main effect of phase on subjective cue weights as well as Cue Weight \times Phase interaction observed in Experiment 2 (which was not found in Experiment 1) is indicative of robust changes in decision strategy as well as increased decision noise under more severe time pressure, which may account for the significant performance difference between the NP and TP phases reported above. In addition, we observed significant positive correlations between participants' subjective cue weights and ideal weights (see Table 4).

Decision strategy model comparison and selection. In accordance with Experiment 1, in Experiment 2, the optimal cue-integration model (Model 15) had the highest exceedance probability (.92) when participants were under no time pressure (NP phase). Under 500 ms time pressure (TP phase), however, Model 5, using only the two most predictive cues, c_1 and c_2 , was the most likely strategy model, with an exceedance probability of .94 (Figure 4D). The control analysis using informative priors also revealed the same pattern of strategy shift (Figure 4F). Note that the Bayesian model comparison enables us to distinguish between whether the least valuable cue was assigned a small weight (perhaps as a consequence of generally lower cue weights in the speeded conditions) versus it being simply ignored altogether, and our results support the latter strategy. Again, switching from the optimal model to Model 5 results in reasonably high ("good-enough") expected accuracy (87%) but with only half of the number of cues to consider, which demonstrates the adaptive nature of cue usage under severe time pressure (Figure 4B).

As in Experiment 1, we next sought to test whether the dropping of the two least valuable cue dimensions was strategic in nature, or whether participants were still trying to employ these cues but ran out of time. Accordingly, we analyzed performance in the TP phase for trials where the two most important cues were identical between the two stimuli (32 trials/phase). Average RT for these trials was 369 ms, which was well below the response deadline of 500 ms (one-sample t test: $t(35) = 61.5, p < .001$), and thus argues against the running-out-of-time hypothesis. Next, as for Experiment 1, we performed a set of analyses on the choice data to determine whether participants performed at chance-level on these trials in the TP phase. Here, 33 out of 36 participants (92%) performed at chance level (binomial test, $p > .05$). Additional analyses based on BFs revealed that 26 out of 36 participants (72%) had a $BF > 1$ (favoring the hypothesis of chance-level performance), with 17 participants showing a significant effect ($BF > 3$). Next, we used BF analysis for testing for significant differences in performance between NP and TP phases using a paired t test (Rouder et al., 2009). This analysis yielded a significant difference between the NP ($mean = 59.5\%$, $SEM = 2.4$) and TP ($mean = 49.3\%$, $SEM = 2.0$) phase performance ($BF = 67.4$, favoring alternative), and importantly, a one-sample t test of the TP phase performance showed a $BF = 5.2$, favoring the null hypothesis, that is, significant evidence for group performance being at chance-level in this condition. Thus, unlike in Experiment 1, in Experiment 2, we obtained unequivocal support for the hypothesis that the dropping of the weakest cues was strategic and categorical, rather than a function of participants running out of time to employ these cues. The reason for this significant finding in Experiment 2 compared to equivocal support in Experiment 1 might of course be the more severe degree of time pressure, but it might also be partly attributable to the fact that this analysis entailed a higher number of trials in Experiment 2 (32 per phase per participant, as compared with 16 in Experiment 1).

To address the apparent discrepancy between the results of subjective cue weights and strategy model selection in the TP phase, as in Experiment 1 we here repeated the analysis using only the high performing participants (average performance of the learning phase $\geq 65\%$). Given that the magnitude of cue weights are highly dependent on individual task performance, a significant decrease in the percentage of optimal responses along with the increase in number of no-response trials in Experiment 2 may have contributed to the overall down-weighting of estimated subjective cue weights, which was not observed in Experiment 1. The rerun of the Bayesian model selection analysis using only the least noisy half of our data set revealed the equivalent results as we obtained above (data not shown). This suggests that the overall decrease of subjective cue weight estimates in the TP phase reflects an increased decision noise and a lower number of observations. The results of Bayesian model selection analysis, indicating a dropping of the weakest cues from the decision process, however, does not appear to be driven by this increase in noise.

In sum, the decision strategy analysis in Experiment 2 documented again that participants engaged in optimal cue integration when time pressure was low, but when faced with time pressure, they satisficed through cue discounting; dropping some information sources from the decision-making process. Importantly, the severe time pressure (500 ms) applied in Experiment 2 led participants to disregard the *two* least predictive cues, compared with the

dropping of only the single least predictive cue under moderate pressure (750 ms) in Experiment 1.

Survey results. When asked about the relative importance of cue states within each cue dimension, the majority of participants were able to correctly identify the higher cue states with a mean correct response of 3.3 ($SD = .7$) out of 4. There was no significant difference in these results between Experiment 1 and 2 ($t(77) = .7, p = .49$), indicating that regardless of experimental group, participants were able to explicitly identify the more predictive cue states in each cue dimension to a similar extent. Similar to Experiment 1, participants' ratings of relative rankings of cue dimensions showed a significant positive correlation with the subjective cue weights and the ideal weights, although this association was more modest for the ideal weights (see Table 4).

When asked about the number of cue dimensions used during each phase, participants indicated that they used significantly fewer cues ($t(35) = 6.2, p < .001$) during the TP phase ($mean = 1.6, SD = .7$) compared with the NP phase ($mean = 2.5, SD = .7$). These results, however, were not significantly different from those of Experiment 1 for both the NP phase ($t(77) = .37, p = .71$) and the TP phase ($t(77) = 1.28, p = .21$). In addition, there were only a total of four (out of 36) participants, whose indicated subjective strategy models matched the objectively identified strategy models for both experimental phases. Thus, as in Experiment 1, we observed some evidence for explicit knowledge of cue values, but little evidence that participants had insight into their decision-making strategies.

Discussion

In Experiment 2, we successfully replicated and extended the results of Experiment 1. First, participants again displayed reliable statistical learning of the probabilistic cue values, and engaged in exhaustive cue integration when solving the task in the absence of time pressure (NP phase). Second, we obtained stronger evidence for the use of the "drop-the-worst" satisficing strategy, in that (a) increasing time pressure from moderate (Experiment 1) to severe (Experiment 2) led to a further trimming of cue usage (from the single to the two least valuable cues); and (b) in selectively analyzing trials where the two highest valued cues were identical between stimuli, we observed clear evidence for intentional ignoring of the weakest cues, and against the notion that participants simply ran out of time in trying to integrate all available cues. In sum, participants strategically disregarded the two least predictive cues to adapt to severe time pressure, although there was little evidence of explicit knowledge about this shift in cue usage in the survey data.

Experiment 3: Satisficing Using Noncompound Cues Under Time Pressure

Many of the previous studies that documented the use of lexicographic heuristics under satisficing pressures have used noncompound stimuli (e.g., Bröder, 2000; Dieckmann & Rieskamp, 2007; Payne, Bettman, & Johnson, 1988; Rieskamp & Otto, 2006), presenting cues independently, for example in the form of an information matrix. In Experiments 1 and 2, we took a departure from this approach and used compound cue stimuli, which integrated all cue dimensions to construct a single object. This raises

the question whether our findings of a drop-the-worst satisficing strategy is for some reason unique to the case of integrated, compound cues. To investigate the generalizability of our “drop-the-worst” findings to noncompound cues, we therefore conducted a third experiment, where we adopted the same type of probabilistic classification task that was used in the previous experiments but using noncompound cue presentation. Given that it is difficult to create unambiguous segregated cues out of some of the cue dimensions we employed in Experiments 1 and 2 (e.g., we cannot show a contour stimulus that does not also have a shape), we here opted to use a new set of cue symbols that lend themselves well to being presented in a cue matrix.

Method

Participants. A new group of 48 volunteers was recruited through a separate HIT on Amazon MTurk with the goal of collecting 38 usable data sets. The same HIT approval criteria used in Experiments 1 and 2 were applied to ensure high data quality. Data from nine participants were excluded, one because of data loss, and eight due to chance-level performance (one-tailed binomial test on percent optimal responses in the initial 480 trials without time pressure; $p > .05$), leaving a total of 39 participants (mean age = 33.9 years, $SD = 8.2$, 21–55 years; 23 female, 16 male). Participants were compensated with \$4.00 upon completion of the experiment and an additional bonus ranging from \$0.50 to \$2.00 based on their performance.

Stimuli. The task stimuli consisted of four pairs of unique cues that comprised each of four cue dimensions (see Figure 5). Similar to the previous experiments, each cue dimension was binary and consisted of two cue states that belong to the same category. For instance, the cue dimension of “weather” had two possible states, sunny and rainy. In addition, each cue dimension was assigned to a fixed location in a stimulus composed of a 2×2 matrix, so that the weather cue always appeared in the top left (location 1), the transportation cue in the top right (location 2), the activity cue in the bottom right (location 3), and the building cue in the bottom left (location 4; see Figure 5A). The same weights as in Experiments 1 and 2 (see Table 2), were randomly assigned to the different cue dimensions for each participant at the beginning of the experiment. Hence, except for the use of noncompound cues, the cue weight assignment and probabilistic nature of the stimuli remained identical to Experiments 1 and 2.

Task. Participants performed a probabilistic classification task, in which they compared two stimuli each composed of four cues and chose the one that is most likely to win (Figure 5B). Stimuli (194×194 pixels each) were presented on the left and the right side of the screen (window size: $1,000 \times 700$ pixels) along the horizontal meridian, at an eccentricity of 250 pixels from a central fixation cross. Similar to the previous experiments, by exhaustively combining all possible cue states, we constructed 120 unique trials, in which the stimuli could differ in one to four cue dimensions. Stimuli were presented on the screen until a response was made or for the duration of an assigned response window of 3 s or 750 ms. Upon each valid response, probabilistic feedback (determined by Equation 1), indicating “win” or “lose,” was displayed for 1 s, followed by a 1-s intertrial interval. Participants were provided with identical task instructions as Experiment 1 (adapted to the new cue symbols), and earned 1 point for every

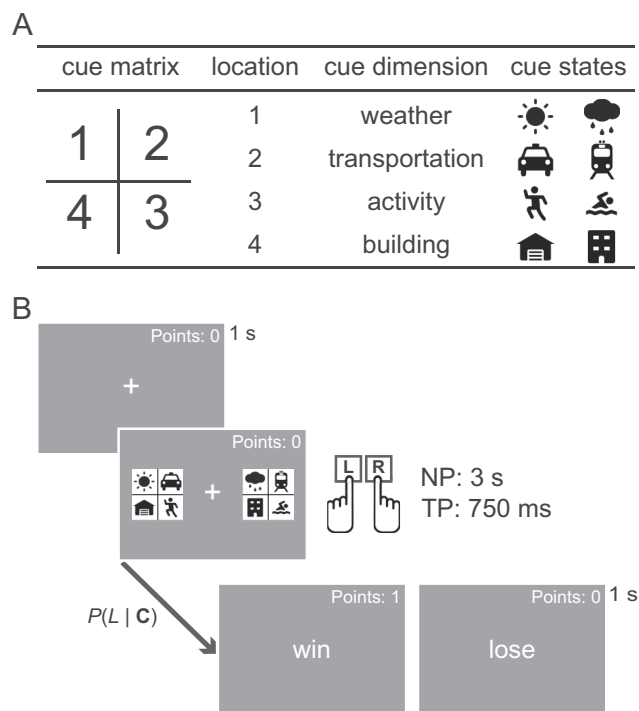


Figure 5. Cue organization and task of Experiment 3. (A) We constructed 16 unique noncompound stimuli by combining four cue dimensions (weather, transportation, activity, and building), each with binary cue states of varying weights (see Table 2). (B) On each trial, participants performed a probabilistic classification task by selecting the stimulus they deemed more likely to “win” under 3 s (initial learning, and NP phase) and 750 ms (TP phase) response deadline. Identical to Experiments 1 and 2, feedback was determined probabilistically based on Eqn. 1, and displayed for 1 s upon each response. NP = no time pressure; TP = time pressure.

winning trial. The total cumulative score at the end of the experiment was used to determine the amount of a bonus payment for each participant.

Again, to compare decision making strategies between conditions with and without time pressure, participants completed three phases sequentially: (a) an initial learning phase (Set 1 and 2); (b) an NP phase (Set 3 and 4); and (c) a TP phase (Set 5 and 6), each of which consisted of 240 trials grouped into four blocks of 60 trials. In both the initial learning period and the NP phase, participants were given a maximum of 3 s to make a decision. This response window was chosen based on a pilot study, which used a 2-s response window for Phases 1 and 2. In this study, participants achieved an average of 70.6% optimal responses at the end of the NP phase (Set 4), which was significantly lower than the NP phase performance of Experiment 1 ($t(79) = 2.4$, $p = .02$). Therefore, to provide participants with sufficient time to observe the cues and thereby to facilitate the learning process, we extended the response window to 3 s. Following the NP phase, participants performed an additional 240 trials of the TP phase with 750 ms time pressure. Identical to the previous experiments, probabilistic feedback was provided throughout the entire experiment.

Survey. At the end of the experiment, participants completed a brief survey, which tested participants’ explicit knowledge of cue

values and their decision strategies in each phase. All the questions were identical to the previous version used in Experiments 1 and 2, except that the choices of cue dimensions and states were modified to match noncompound stimulus conditions. Based on the answers to the survey questions, self-report strategy models were constructed.

Data analysis. Data analyses were carried out identical to Experiments 1 and 2.

Results

Task performance. Similar to Experiments 1 and 2, participants were able to gradually improve their performance throughout the initial learning phase (Figure 6, Table 3). Even after the completion of the first 240 trials, participants continued to learn, indicated by significant difference in the percentage of optimal choices between Set 2 and 3 ($t(38) = 3.2, p < .01$). We found no difference in performance between Set 3 and 4 ($t(38) = 1.4, p = .16$), suggesting that performance had stabilized during the NP phase. The percentage of optimal choices and RT were influenced by decision difficulty, which is characterized by the difference in sum of cue weights between the two stimuli, with more difficult decisions associated with decreasing performance (linear trend: $F(1, 38) = 269.5, p < .001$) and increasing RT (linear trend: $F(1, 38) = 43.4, p < .001$).

In the subsequent TP phase (see Figure 6), participants experienced a time pressure of 750 ms, which was substantially below the mean RT of the NP phase (see Table 3). Accordingly, both the rate of optimal choices ($t(38) = 8.9, p < .001$) and RT ($t(38) = 12.4, p < .001$) decreased significantly compared with the NP phase. We found no significant difference in performance between Set 5 and 6 ($t(38) = .3, p = .74$). Regardless, the effect of decision difficulty on the percentage of optimal choices (linear trend: $F(1, 38) = 134.9, p < .001$) as well as RT (linear trend: $F(1, 38) = 4.2, p < .05$) was still present in the TP phase. Similar to the previous experiments, there was no significant effect of gender on performance (NP: $t(37) = 1.8, p = .08$; TP: $t(37) = 1.7, p = .10$), and no significant correlation between age and performance (Pearson's

correlation, NP: $r = .1, p = .67$; TP: $r = -.1, p = .77$). Taken together, using noncompound stimuli, we obtained a comparable statistical learning profile to that observed in Experiments 1 and 2. That is, participants successfully learned to solve the task through trial-and-error learning over the course of the initial learning and the NP phases. In the TP phase, however, time pressure had a detrimental effect on performance, thus creating conditions for possible changes in decision strategy.

Subjective cue weights. To examine the effect of each cue dimension on participants' decisions, we again carried out logistic regression based on choice data in the NP and TP phases (Equations 2–4, Figure 7A). In accordance with our previous findings based on compound stimuli, an ANOVA with assigned cue weights (w_i) and experimental phase showed a significant main effect of cue weights ($F(2.0, 75.5) = 26.8, p < .001$), which was characterized by a significant linear trend ($F(1, 38) = 73.8, p < .001$). This suggests that participants had correctly learned the relative importance of the different cues even with noncompound stimuli. In addition, there was a significant effect of NP/TP phase ($F(1, 38) = 48.0, p < .001$) with marginal Cue Weight \times Phase interaction ($F(3, 114) = 2.6, p = .06$). Post hoc test revealed that there was a significant overall down-weighting of cues in the TP phase compared to the NP phase (mean difference = .23, $SE = .03, p < .001$). This main effect of experimental phase, as in the case of Experiment 2, is indicative of a shift in decision strategy along with increased decision noise under time pressure, which also accounts for a significant decrease in performance in the TP phase compared with the NP phase. Moreover, comparison between subjective cue weights and weights produced by an ideal observer revealed a significant positive correlation (see Table 4), which further confirms that, on average, participants learned to accurately rank the cue dimensions.

Decision strategy model comparison and selection. We again used variational Bayesian inference (Drugowitsch, 2013) and Bayesian model comparison (Stephan et al., 2009) to quantify participants' decision making strategies at an individual as well as at the group level (see Method in Experiment 1). In the NP phase, unlike Experiments 1 and 2, the model selection resulted in no clear winning model, with exceedance probability evenly shared among Model 15 (.37 for uninformative and .23 for uninformative priors), Model 11 (.33 for uninformative and .28 for uninformative priors), and Model 5 (.28 for uninformative and .47 for informative priors; Figure 7B, C; see also Figure 4A). This suggests that unlike the comparison between compound stimuli, when subjects are comparing noncompound stimuli, they are more inclined to consider a lower number of cues even when the time pressure is fairly low. It is, however, worth noting that this reduction of cue usage appears to be strategic and adaptive, and conforms to the drop-the-worst strategy: compared to the optimal model (Model 15), Model 11 ignores the least important cue with 97% expected accuracy, and Model 5 ignores the two least important cues with 87% expected accuracy, both of which result in only small drop-offs in performance (Figure 4B). By contrast to the NP phase, performance in the TP phase, with a 750-ms response window, was best captured by strategy Model 1. This model, which only uses the most valuable cue dimension and thus ignores the three weakest cues, was the most likely strategy model with an exceedance probability of .85

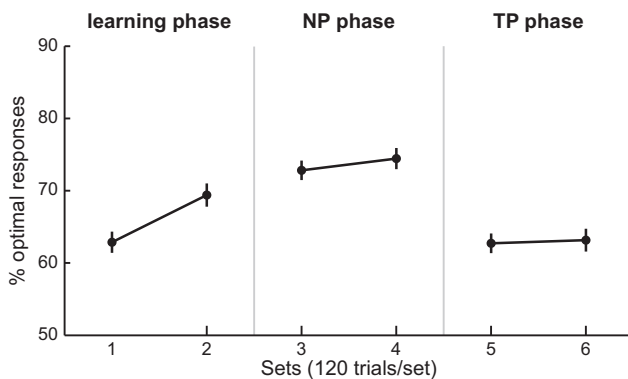


Figure 6. General task performance of Experiment 3. Average percent optimal (correct) responses in the learning (left), NP (3-s response window; middle) and TP (750-ms response window; right) phases. An optimal choice refers to the “correct” decision favored by the sum of cue weights, independent of the actual feedback. Trials with two stimuli with equal sum of weights were excluded. Error bars are standard errors (SEM). NP = no time pressure; TP = time pressure.

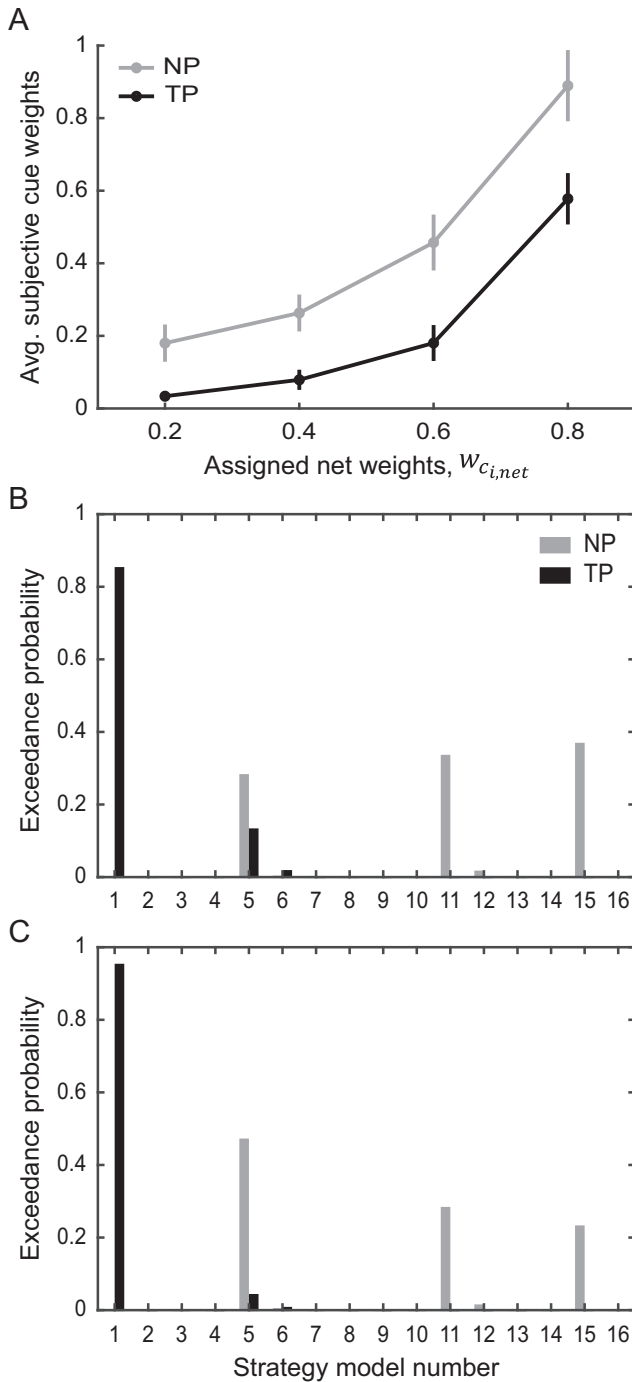


Figure 7. Subjective cue weights and Bayesian strategy model selection. (A) Average subjective cue weights as a function of NP (gray line) and TP (black line) phases for Experiment 3. Error bars are standard errors (SEM). (B-C) Bayesian strategy model selection group results for the NP (gray) and TP (black) phases of Experiment 3 using (B) uninformative priors and (C) informative priors. NP = no time pressure; TP = time pressure.

(uninformative priors) and .95 (informative priors, Figure 7B, 7C). These results conceptually replicate those of Experiments 1 and 2, in that they suggest that under higher time pressure participants reduced the search space following a drop-the-

worst principle, here using only the single highest cue, c_1 , to arrive at their decisions.

Next, as in Experiments 1 and 2, we aimed to tackle the question whether the observed shift in strategy from the NP to TP phase is strategic in nature or participants simply didn't have enough time to process additional cues. Accordingly, we analyzed performance in trials when the most important cue was identical between the two stimuli (112 trials/phase). Average RT for these trials in the TP phase was significantly below the response deadline of 750 ms ($mean = 519$ ms; one-sample t test: $t(38) = 18.6, p < .001$), which is in contradiction with the running-out-of-time hypothesis. To further examine the validity of the model selection result, we conducted a set of additional analyses to test whether performance on these trials is at chance in the TP phase. This analysis revealed that 35 out of 39 participants (90%) performed no different from chance level (binomial test, $p > .05$), indicating that the great majority of participants guessed when the most important cue did not discriminate between the two stimuli. Additional analysis based on BFs revealed that 35 out of 39 participants (90%) had a $BF > 1$ (favoring the hypothesis of chance-level performance), with 28 participants showing a significant effect ($BF > 3$). At the group level, we performed a Bayesian paired t test (Rouder et al., 2009) to compare average performance between the NP ($mean = 64.7\%$, $SEM = 1.7$) and TP ($mean = 51.5\%$, $SEM = 1.1$) phases. This t test yielded a highly significant difference ($BF = 4.4 \times 10^7$, favoring alternative), and an additional one-sample t test of the TP phase performance revealed $BF = 2.5$ (standard t test: $t(38) = 1.3, p = .19$), thus favoring the null hypothesis that group performance is at chance level on these trials under time pressure. In sum, using noncompound cues, we again obtained evidence supporting that dropping of the weakest cues is a strategic decision to cope with time pressure.

In summary, the decision strategy analysis in Experiment 3 revealed that when multiple cues were presented in a spatially segregated manner, participants were more prone to consider a lower number of cues and to adopt a noncompensatory strategy even without high time pressure. Under 750 ms time pressure, participants satisficed by dropping the weakest three cues from their decision making process, considering only the highest cue and resorting to guessing when this cue did not discriminate between the two cue matrices. In addition, Bayesian model comparison showed that only three participants out of 39 (one in the NP and two in the TP phase) preferred the take-the-best model (Model 16) over the optimal cue integration model, suggesting that the majority of participants did not adopt a sequential, lexicographic strategy. Rather, participants seem to strategically disregard the least predictive cues and focus only on a small subset of cues to guide their decisions.

Survey results. Similar to Experiments 1 and 2, most participants were able to correctly identify the cue states with a higher predictive value within each cue dimension with a mean correct response of 3.7 ($SD = .6$) out of 4. Participants' explicit ratings of relative importance of cue dimensions showed a significant positive correlation that matched very closely to their inferred subjective cue weights as well as the ideal weights (see Table 4).

When asked about the number of cue dimension usage in each phase, participants indicated that they used significantly fewer cues ($t(38) = 12.5, p < .001$) in the TP phase ($mean = 1.5, SD = .5$) compared with the NP phase ($mean = 2.8, SD = .7$). Com-

parison between subjective strategy models, which were defined by each participant's response to the survey questionnaire, and objectively identified strategy models revealed that only a total of five out of 39 participants were able to correctly identify their strategies for both experimental phases. Thus, in accordance with our previous findings, we obtained some evidence for explicit knowledge of relative importance of cue dimensions and states, but only ambiguous support for any insight into the decision-making strategies employed in the NP and TP phases.

Discussion

In Experiment 3, we sought to examine whether our findings of a “drop-the-worst” satisficing strategy based on compound cues can be generalized to a probabilistic classification task using noncompound cues. We found that similar to the previous experiments, participants were able to learn the probabilistic cue values through trial-and-error. In the NP phase, however, only a small portion of participants appeared to engage in exhaustive cue integration and many participants adopted suboptimal strategies. Importantly, though, those suboptimal strategies conformed to a drop-the-worst pattern, with participants using only the two or three highest value cues to make decisions. In the 750 ms TP phase, participants further reduced the search space and considered only the single most important cue. Our model selection results provided no support for widespread use of the take-the-best strategy. In sum, using noncompound cues, we successfully replicated our main results from Experiments 1 and 2, confirming the reliable use of the “drop-the-worst” strategy.

The difference in the results of the NP phase from our previous findings may have been caused by participants' strategic decision to decrease the number of cues to consider to reduce their effort in the presence of low, but not negligible, time pressure. As reported in Table 3, participants took longer to arrive at decisions in the learning and NP phases in Experiment 3 compared with the experiments using compound cues, suggesting that comparing noncompound multicue stimuli may be more effortful than comparing compound stimuli. In addition, previous studies that reported the use of the weighted additive strategy using noncompound cues have provided participants with ample time to observe and compare the cues before making a choice (e.g., Glöckner & Betsch, 2008; Pachur & Olsson, 2012). However, our task imposed a fixed response deadline of 3 s during the learning and NP phases, which may have already encouraged the use of suboptimal strategies. Regardless, rather than adopting the lexicographic, take-the-best heuristic that relies on a single discriminating cue through sequential search (Gigerenzer & Goldstein, 1996; e.g., Payne, Bettman, & Johnson, 1988; Rieskamp & Hoffrage, 2008), or dropping random cues from their decision process, participants systematically integrated the two or three most valuable cues to arrive at decisions in the NP phase.

In the TP phase, participants strongly relied on the single highest cue to solve the task. Through a set of control analyses, we were able to confirm that participants performed at chance when faced with trials when the highest cue did not discriminate between two stimuli. It is, however, difficult to clearly distinguish whether participants were trying to employ the take-the-best strategy but did not have enough time to evaluate the second highest cue or they strategically disregarded the rest of the cues altogether. Given

the saliency of the cues as well as RT data, it is unlikely that participants did not have enough time to process the second cue visually but it is difficult to completely rule out the possibility of the use of lexicographic strategy under high time pressure.

General Discussion

Decision-making in everyday life is beset by uncertainty due to noisy and incomplete information, and limited available decision time and cognitive resources. It has long been held that humans adaptively use satisficing strategies that can simplify the decision-making process to save time and cognitive effort to arrive at good-enough solutions (Gigerenzer & Goldstein, 1996; Shah & Oppenheimer, 2008; Simon, 1955, 1956, 1990), and some previous studies have begun to investigate the nature of shifts in decision-making under time pressure. The goal of the present study was to expand this literature by creating a protocol that emphasized two key aspects of real-life, high-stakes decision making, namely uncertainty of cue values combined with severe time constraints, that is, split-second decision making. Specifically, we examined how people solve a probabilistic classification task under varying time pressure, thereby connecting the often disparate literatures on learning and decision making. Using compound cues, we found that, under low time pressure (NP phase), participants were able to correctly weight and integrate all available cues to arrive at near-optimal decisions. With increasing time pressure (TP phase), however, participants shifted their decision strategies by dropping cues from the information-integration process. Importantly, this selective discounting of a subset of cue information was clearly strategic and adaptive, in that participants specifically dropped the one (Experiment 1, moderate time pressure group with 750-ms response window) or two (Experiment 2, severe time pressure group with 500-ms response window) least informative cues from the decision-making process. Moreover, control analyses confirmed that disregarding of the least valuable cue(s) was not an expression of simply running out of time during an attempt at integrating those cue values. Rather, the weakest cues seem to have been categorically excluded from the decision process under high time pressure. We replicated these results using noncompound cues (Experiment 3), which demonstrated that under 750-ms time pressure, participants dropped the three least informative cues, utilizing only the best cue to make decisions. Postexperiment survey results suggested that participants had at least some explicit knowledge concerning the informative value of the cues but lacked insight into the decision strategies they adopted. Our results thus document, and quantify, adaptive shifts in decision strategies under uncertainty to compensate for limited decision time. Specifically, we showed that participants engaged in adaptive cue discounting, ignoring the least valuable information sources, a satisficing variant we here call “drop-the-worst.”

Our discovery of the decision strategy that drives satisficing under uncertainty and high time-pressure is significant, because this knowledge, in principle, renders split-second human choices (e.g., in traffic or combat) predictable, which in turn can inform the optimal design of safety measures and/or autonomous agents that interact with humans. The fact that the nature of satisficing in split-second decision making cannot be anticipated on the basis of strategies uncovered in more slow-paced environments becomes clear when we contrast the present findings with those of several

prior investigations using cue-based search paradigms. These studies examined the effect of time pressure on strategy selection in slow-paced (>15 s–50 s) scenarios and have found that under low time pressure, similar to the present study, a strategy model that integrates all available information (e.g., weighted additive strategy) performed best at predicting participants' choices (Rieskamp & Hoffrage, 2008). However, under higher time pressure, people adapted by using a simple lexicographic heuristic (i.e., a "take-the-best" heuristic for a binary choice problem; Gigerenzer & Goldstein, 1996; Payne et al., 1988, 1996; Rieskamp & Hoffrage, 2008), which looks for "one clever cue" to base decisions on. These conclusions are partly drawn from the information search structure of tasks that allow subjects to inspect only one piece of information at a time (Payne et al., 1988; see Glöckner & Betsch, 2008 for discussion). Specifically, take-the-best requires searching through cues in descending order of cue validity until the decision maker finds the first (and the "best") cue that discriminates between the alternatives and such sequential search paradigm can strongly encourage subjects to adopt one-cue heuristic under time pressure. Indeed, when cues were presented all at once, people were able to integrate cue information using compensatory strategies relatively quickly (Glöckner & Betsch, 2008; Pachur & Olsson, 2012), which is in accordance with our findings under no time pressure. Our present findings further extend the previous work on bounded rationality by documenting, for the first time, that under much higher time pressure that enforces split-second decision making, subjects instead adopted a "drop-the-worst" heuristic, whereby the least valuable cues are simply ignored (or discounted) altogether.

This difference in satisficing strategy likely represents an adaptive shift in information integration to accommodate the less certain nature of cue information and much higher time pressure in the present task. First, in order to implement the take-the-best approach, it is crucial for the decision maker to be confident in their knowledge of the exact rankings of cue validities; in fact, the majority of previous work on this heuristic has provided participants with explicit cue validities on each trial (e.g., Bröder, 2000, 2003; Dieckmann & Rieskamp, 2007; Payne et al., 1988, 1996; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006). Moreover, it has been shown that experts who have better knowledge of cue validities than novices are more likely to adopt the take-the-best heuristic (Garcia-Retamero & Dhami, 2009; Pachur & Marinello, 2013). By contrast, participants in our study were not provided with explicit cue weights and had to infer the correct cue ranking through trial-and-error learning, likely rendering them less confident about the exact rankings. In addition, due to a large number of possible cue combinations, each associated with probabilistic feedback, our experimental design likely prevented subjects from making memory-guided decisions (see Juslin, Olsson, & Olsson, 2003). Given this uncertainty about cue information and the complexity of the task structure, it may be more adaptive to first set a satisfactory cutoff level, integrate cues in an order that is most likely to reach this cutoff, and then make a decision without evaluating all available cues (Shah & Oppenheimer, 2008), especially under high time pressure. This strategy reduces cognitive effort by integrating less information and choosing an alternative that is simply "good-enough" whereas take-the-best reduces effort by examining only one cue at a time, but it can require searching

for multiple cues if the first cue does not discriminate between alternatives (Shah & Oppenheimer, 2008).

Second, given that take-the-best requires serially searching through the cues in order of their validity until one finds the first discriminating cue, in some cases this strategy can require more time than the drop-the-worst strategy, in particular when the least informative cue (Experiment 1) or cues (Experiments 2 and 3) are the only cues differentiating between the two stimuli. Therefore, if there is high pressure to make a quick decision with some uncertainties in cue validities, lowering the satisfactory cutoff level and hence, focusing only on a fixed subset of cues in the decision process may be a more efficient way to save cognitive effort and achieve good-enough accuracy. Accordingly, our findings document that participants did not employ the take-the-best strategy (Model 16), even when the cues were presented separately, but instead used a drop-the-worst approach, where only a subset of cues with the highest validities (weights) was considered, and random guesses were made if those cues did not differ between the two stimuli. Although Bergert and Nosofsky's (2007) response-time approach demonstrated more use of take-the-best strategy compared with the weighted additive strategy using compound cues, the use of deterministic feedback during the training phase as well as a task structure that yielded comparable results between the use of take-the-best and the optimal strategies may have influenced the participants to rely on a single cue heuristics (see Juslin et al., 2003). It is also important to note that the observed strategy shift to noncompensatory heuristics can be induced by the effect of learning and experience over the course of the experiment (Garcia-Retamero & Dhami, 2009; Johnson & Payne, 1986; Pachur & Marinello, 2013; Rieskamp & Otto, 2006). Our results are no exception to this observation, and because participants were provided with probabilistic feedback throughout all phases of the experiment, and the TP phase always followed the NP phase, it is possible that continued learning may also have played a role in inducing the switch to the drop-the-worst strategy.

One intriguing question arising from the present protocol is whether multiple cues are processed in a serial manner (i.e., integrating cues serially from the most important to the least using an additive rule) or in a parallel fashion. For instance, one could imagine evidence for all cue dimensions being accumulated simultaneously in a drift-diffusion type model (Smith & Ratcliff, 2004), where the drift rate of evidence may increase as a function of cue weight, corresponding to an effect of attention. With our current paradigm, however, it is difficult to distinguish whether behavior stems from a serial or parallel cue processing strategy, because equivalent predictions seem to follow from both models. For instance, in the NP phase of both experiments, we found significant linear trends of RT as a function of cue weights when only a single cue was different between two stimuli (Experiment 1: $F(1, 42) = 22.1, p < .001$; Experiment 2: $F(1, 35) = 4.1, p < .05$). In other words, participants took longer time to make a choice when a less valuable cue was the only distinguishing feature between the two stimuli, in spite of the fact that this single difference should be perceptually quite salient. This might be interpreted as support for the serial processing model, as participants may have evaluated cues serially in order of importance, thus leading to slower response times for less valuable cues even when only that cue distinguishes the two stimuli. However, this result

could equally be driven by decision difficulty, with more difficult trials (i.e., with less cue value difference between stimuli) requiring more time to make a choice. Taken together, the precise manner in which multiple cues are integrated and how the shift in decision strategy is instantiated computationally and neurally under time pressure represents an exciting challenge for future research.

Several previous studies have explored ways of identifying strategies that participants use in solving the classic weather prediction task (Gluck, Shohamy, & Myers, 2002; Lagnado, Newell, Kahan, & Shanks, 2006; Meeter, Myers, Shohamy, Hopkins, & Gluck, 2006; Meeter, Radics, Myers, Gluck, & Hopkins, 2008; Speekenbrink, Lagnado, Wilkinson, Jahanshahi, & Shanks, 2010), which has close parallels with the present protocol. Gluck et al. (2002) introduced a method using a least-mean-squared-error measure that compares each participant's data to the ideal response profiles constructed from three different strategy models, the multicue (optimal model), one cue, and singleton strategies. Hence, a strategy that resulted in the lowest error was defined to be the best-fit model. This model-based approach was later extended by using Monte Carlo simulations that can be harnessed to infer switches from one strategy to another over the course of the experiment (Meeter et al., 2006). Another approach in identifying an individual's strategy is to use "rolling regression" methods that estimate subjective weights of each cue through a moving window of consecutive trials, which can be applied to characterize how the learning occurs during the course of the task (Kelley & Friedman, 2002; Lagnado et al., 2006). Meeter et al. (2008) demonstrated that both strategy analysis and rolling regression analysis results in a more or less equivalent ability to predict responses.

We here applied a new analytical approach that captures these key aspects of prior strategy analyses and enabled us to infer both subjective cue values and the manner in which these were combined to reach decisions. Specifically, we inferred how the cues were weighted in each phase of the experiment using logistic regression, and identified the most likely strategy employed by using Bayesian model comparison at an individual as well as the group level. Hence, instead of simply categorizing strategies into the number of cues used (cf. Gluck et al., 2002; Lagnado et al., 2006; Meeter et al., 2008), we explored an exhaustive set of plausible strategy models to identify the exact cues used and their relative importance in making decisions. For instance, even if a given participant adopted a suboptimal strategy that resulted in poor performance, we were nevertheless able to infer the most likely underlying cue structure that the participant may have developed throughout the task. The present analysis approach therefore may have great potential for enhancing the inferences drawn from future studies of statistical learning.

Moreover, the present results, in accordance with previous studies using the classic weather prediction task (Gluck et al., 2002; Lagnado et al., 2006; Meeter et al., 2006, 2008; Speekenbrink et al., 2010), highlight the fact that there is likely considerable variability in participants' cognitive strategies in probabilistic decision making scenarios. Previous studies reported that people tend to start with a simple strategy using a single cue (i.e., singleton strategy) but switch to an optimal

multicue strategy toward the end of the task. Although we did not assess changes in strategies across multiple time points within each experimental phase (which would require a yet higher trial count), our performance data suggest that participants learned to use the cues in a near-optimal manner as they learned the cue characteristics throughout the initial learning and the NP phase (see Figure 4B, 4C). In addition, our data on the subjective cue weights indicate that the majority of participants learned the correct relative importance of cue weights.

Another intriguing model for solving a multicue probabilistic classification task, especially under high time pressure, is a mixture strategy model where participants some of the time integrate cues in an optimal manner and some of the time decide randomly. We had considered the possibility of this type of mixture model, but rejected it based on the fact that this model predicts psychometric curves that do not saturate to 100% accuracy when the sum of evidence is at the maximum value, unless when the guessing parameter is set to zero. By contrast, the suboptimal cue usage models such as Models 5 and 11, predict that accuracy will suffer more when the evidence is low but the performance will eventually saturate toward 100%. Note that this pattern also holds for predictions at very low guess rates ($g \leq .2$), which are in a reasonable range if we consider decision noise and imperfect knowledge of cue weights. Hence, even though participants are making more errors (possibly random guesses) under time pressure, the performance of a majority of participants, especially the high performers, is scaled by the sum of evidence, reaching the peak when the evidence is at maximum. The shape of these performance curves corresponds to the mixture model with $g \leq .2$, indicating that performance in the TP phase does not reflect a random increase in the guess rate, but is reflective of a systematic lapse in decision making dependent on the sum of evidence. In addition, in case of severe time pressure (Experiments 2 for compound cues and Experiment 3 for noncompound cues), we showed that performance is at chance when the low-value cues are the only informative cues. Therefore, although the mixture model can account for some proportion of our observations, because our goal was to uncover patterns of cue usage under varying degrees of time pressure, we did not include the model in our strategy model comparison.

Finally, there have been mixed findings regarding whether participants have explicit knowledge about how they solve the classic weather prediction task (Gluck et al., 2002; Knowlton et al., 1994; Lagnado et al., 2006; Newell, Lagnado, & Shanks, 2007). This knowledge can be divided into two different insights that might not necessarily coincide: (a) insight into the cue structure of the task, and (b) insight into strategy use (Lagnado et al., 2006). Gluck et al. (2002) reported that there was little or no evidence that participants had explicit insight about the task structure or their strategy use based on their postexperiment questionnaire. On the other hand, Lagnado et al. (2006), using more detailed and frequent measures, found that having accurate knowledge of the task structure and self-insight is necessary for achieving optimal performance. Our postexperiment survey results seem to indicate that the majority of subjects in the present study had some explicit knowledge about the cue structure, but that this knowledge did not necessarily lead to having accurate insight into strategy use.

In summary, the present study characterized the nature of adaptive shifts in decision strategies in an uncertain and fast-paced environment. By combining a probabilistic classification task under varying time pressure with new analytical approaches to quantifying decision models, we showed for the first time that, when forced to make split-second decisions in an uncertain environment, participants strategically discount the least valuable piece(s) of information, providing novel evidence for a “drop-the-worst” cue satisficing decision-making strategy.

References

- Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception & Psychophysics*, *73*, 640–657. <http://dx.doi.org/10.3758/s13414-010-0049-7>
- Ballew, C. C., II, & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 17948–17953. <http://dx.doi.org/10.1073/pnas.0705435104>
- Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 107–129. <http://dx.doi.org/10.1037/0278-7393.33.1.107>
- Böckenholt, U., & Kroeger, K. (1993). The effect of time pressure in multiattribute binary choice tasks. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 195–214). Boston, MA: Springer. http://dx.doi.org/10.1007/978-1-4757-6846-6_14
- Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, *63*, 863–891. <http://dx.doi.org/10.1080/17470210903091643>
- 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. <http://dx.doi.org/10.1037/0278-7393.26.5.1332>
- 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. <http://dx.doi.org/10.1037/0278-7393.29.4.611>
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*, 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, *35*, 1801–1813. <http://dx.doi.org/10.3758/BF03193511>
- Drugowitsch, J. (2013). *Variational Bayesian inference for linear and logistic regression*. Retrieved from <http://arxiv.org/abs/1310.5438>
- Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *eLife*, *3*, 1–19. <http://dx.doi.org/10.7554/eLife.03005>
- Friston, K., Litvak, V., Oswal, A., Razi, A., Stephan, K., van Wijk, B., . . . Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage*, *128*, 413–431. <http://dx.doi.org/10.1016/j.neuroimage.2015.11.015>
- Garcia-Retamero, R., & Dhami, M. K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review*, *16*, 163–169. <http://dx.doi.org/10.3758/PBR.16.1.163>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. <http://dx.doi.org/10.1146/annurev-psych-120709-145346>
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669. <http://dx.doi.org/10.1037/0033-295X.103.4.650>
- Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1055–1075. <http://dx.doi.org/10.1037/0278-7393.34.5.1055>
- Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the “weather prediction” task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*, 408–418. <http://dx.doi.org/10.1101/lm.45202>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. *Psychology of Learning and Motivation*, *18*, 1–47. [http://dx.doi.org/10.1016/S0079-7421\(08\)60358-8](http://dx.doi.org/10.1016/S0079-7421(08)60358-8)
- Johnson, E., & Payne, J. (1986). The decision to commit a crime: An information-processing analysis. In D. B. Cornish & R. V. Clarke (Eds.), *The reasoning criminal: Rational choice perspectives on offending* (pp. 170–185). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4613-8625-4_11
- Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133–156. <http://dx.doi.org/10.1037/0096-3445.132.1.133>
- Juslin, P., & Persson, M. (2002). PROBABILITIES from EXemplars (PROBEX): A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563–607. http://dx.doi.org/10.1207/s15516709cog2605_2
- Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795. <http://dx.doi.org/10.1080/01621459.1995.10476572>
- Kelley, H., & Friedman, D. (2002). Learning to forecast price. *Economic Inquiry*, *40*, 556–573. <http://dx.doi.org/10.1093/ei/40.4.556>
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402. <http://dx.doi.org/10.1126/science.273.5280.1399>
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*, 106–120.
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*, 162–183. <http://dx.doi.org/10.1037/0096-3445.135.2.162>
- Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*, 161–180. <http://dx.doi.org/10.1037/0096-3445.124.2.161>
- Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*, 227–260. <http://dx.doi.org/10.1037/0033-295X.107.2.227>
- 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. <http://dx.doi.org/10.3758/BF03196581>
- Maule, A. J. (1994). A componential investigation of the relation between structural modelling and cognitive accounts of human judgement. *Acta Psychologica*, *87*, 199–216. [http://dx.doi.org/10.1016/0001-6918\(94\)90051-5](http://dx.doi.org/10.1016/0001-6918(94)90051-5)
- Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning & Memory*, *13*, 230–239. <http://dx.doi.org/10.1101/lm.43006>
- Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience and Biobehavioral Reviews*, *32*, 237–248. <http://dx.doi.org/10.1016/j.neubiorev.2007.11.001>
- Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psycho-*

- nomic Bulletin & Review*, 14, 505–511. <http://dx.doi.org/10.3758/BF03194098>
- 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. <http://dx.doi.org/10.1037/0278-7393.29.1.53>
- Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 923–935.
- 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. [http://dx.doi.org/10.1016/S0749-5978\(02\)00525-3](http://dx.doi.org/10.1016/S0749-5978(02)00525-3)
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300. <http://dx.doi.org/10.1037/0033-295X.104.2.266>
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002. <http://dx.doi.org/10.1037/0278-7393.32.5.983>
- Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers? *Acta Psychologica*, 144, 97–103. <http://dx.doi.org/10.1016/j.actpsy.2013.05.003>
- Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology*, 65, 207–240. <http://dx.doi.org/10.1016/j.cogpsych.2012.03.003>
- Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology*, 2, 147.
- Payne, J., Bettman, J., & Johnson, E. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534–552. <http://dx.doi.org/10.1037/0278-7393.14.3.534>
- Payne, J., Bettman, J., & Luce, M. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes*, 66, 131–152. <http://dx.doi.org/10.1006/obhd.1996.0044>
- Pohl, R. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 271, 251–271. <http://dx.doi.org/10.1002/bdm.522>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108. <http://dx.doi.org/10.1037/0033-295X.85.2.59>
- Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1355–1370. <http://dx.doi.org/10.1037/0278-7393.32.6.1355>
- Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258–276. <http://dx.doi.org/10.1016/j.actpsy.2007.05.004>
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236. <http://dx.doi.org/10.1037/0096-3445.135.2.207>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. <http://dx.doi.org/10.3758/PBR.16.2.225>
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112, 610–628. <http://dx.doi.org/10.1037/0033-295X.112.3.610>
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, 134, 207–222. <http://dx.doi.org/10.1037/0033-2909.134.2.207>
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1865–1897. <http://dx.doi.org/10.1037/a0016926>
- Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69, 99–118. <http://dx.doi.org/10.2307/1884852>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138. <http://dx.doi.org/10.1037/h0042769>
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1–19. <http://dx.doi.org/10.1146/annurev.ps.41.020190.000245>
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27, 161–168. <http://dx.doi.org/10.1016/j.tins.2004.01.006>
- Speekenbrink, M., Lagnado, D. A., Wilkinson, L., Jahanshahi, M., & Shanks, D. R. (2010). Models of probabilistic category learning in Parkinson’s disease: Strategy use and the effects of L-dopa. *Journal of Mathematical Psychology*, 54, 123–136. <http://dx.doi.org/10.1016/j.jmp.2009.07.004>
- Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, 46, 1004–1017. <http://dx.doi.org/10.1016/j.neuroimage.2009.03.025>
- Svenson, O., & Maule, A. J. (Eds.). (1993). *Time pressure and stress in human judgment and decision making*. Boston, MA: Springer. <http://dx.doi.org/10.1007/978-1-4757-6846-6>
- Vulkan, N. (2000). An economist’s perspective on probability matching. *Journal of Economic Surveys*, 14, 101–118. <http://dx.doi.org/10.1111/1467-6419.00106>
- Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1235–1253. <http://dx.doi.org/10.1037/0278-7393.19.6.1235>
- Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60, 181–192. <http://dx.doi.org/10.1037/0022-3514.60.2.181>
- Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, 447, 1075–1080. <http://dx.doi.org/10.1038/nature05852>

Received August 27, 2015

Revision received March 8, 2016

Accepted March 10, 2016 ■