American Psychological Association

Decision

2325-9965/19/$12.00

2019, Vol. 6, No. 4, 335–368
http://dx.doi.org/10.1037/dec0000105

Understanding the Complexity of Simple Decisions: Modeling Multiple Behaviors and Switching Strategies

Michael D. Lee
University of California, Irvine

Kevin A. Gluck
Wright-Patterson Air Force Base, Ohio

Matthew M. Walsh
The RAND Corporation, Pittsburgh, Pennsylvania

We develop models of strategy use in multiple-cue decision making that have a couple of key capabilities. The first is that they incorporate multiple sources of behavior, using the predictions that strategies make about processes and outcomes as a basis for inferring strategy use. The second is that they allow for people to change strategies multiple times over a sequence of decision trials. The models are implemented as generative probabilistic models, allowing for fully Bayesian inference about the nature of strategy use and the number of strategy switches. To demonstrate the approach and evaluate the models, we consider the standard take-the-best, weighted-additive, and tally strategies, as well as a guessing strategy, and apply them to previously published experimental data that involve decision, search, and verbal report data (Walsh & Gluck, 2016). We find strong evidence that many people switch strategies many times, especially when inference is based on all of the available behavioral data. Our results suggest there is interpretable complexity beneath people's use of simple strategies to make decisions.

Keywords: decision strategies, heuristic decision making, take-the-best, tally, weighted additive

Our everyday lives are filled with situations in which we need to make a simple decision between two options. Perhaps it is date night and there are two good movie options at the local theater. Or perhaps your child's club sport team needs volunteer help from the parents, and you can either manage the clock or work in the concession stand during the next game. Or perhaps you are faced with the common conundrum at contemporary professional events of choosing between two boxed lunch options. In some ways, these choices are about as simple as a decision-making task can be. The stakes are not high, nor is there time pressure. The decision involves a small number of alternatives with well-understood features or cues. This clarity contrasts with many real-world decision-making situations that are characterized by uncertainty about possible alternatives, uncertainty about relevant features, and uncertainty about the values of possible features for possible alternatives (Klayman, 1984; Klein, 2008). In choosing between two lunch boxes, for example, there is no uncertainty about the ingre-

This article was published Online First March 28, 2019.
Michael D. Lee, Department of Cognitive Sciences, University of California, Irvine; Kevin A. Gluck, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio; Matthew M. Walsh, The RAND Corporation, Pittsburgh, Pennsylvania.
This research was supported by the U.S. Air Force Research Laboratory's Cognitive Models and Agents Branch. Michael D. Lee's collaboration was enabled through an appointment to the Oak Ridge Institute for Science and Education (ORISE) Faculty Research Program.

A project page associated with this article, containing code and data, is available on the Open Science Framework (OSF) at https://osf.io/s9u5x/.

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. Air Force, Department of Defense, or the United States Government.

Correspondence concerning this article should be addressed to Michael D. Lee, Department of Cognitive Sciences, 3151 SSPA, University of California, Irvine, CA 92697-5100. E-mail: mdlee@uci.edu

dients of the sandwiches, nor of the contents of the box. The ingredients are likely to be familiar, and opinions about their desirability long established. And neither lunch will be so good (nor so bad) that the choice is of much consequence.

Yet even the simple decision of choosing a boxed lunch invites all sorts of complexity in human behavior. There are different possible approaches for finding out about the contents of the box. An exhaustive search finds out what the sandwich ingredients are and what else the box contains. A limited search might only find out about one or two key ingredients in the sandwich. There are also different possible mechanisms for making a choice. A compensatory choice, which considers the available information, usually gives significant weight to a person's preferences for each ingredient. A noncompensatory choice, which considers only some of the available information, might allow one ingredient, such as any sort of meat for a vegetarian, to dictate the choice. Understanding how people choose between lunch boxes becomes even more complicated when choice is expanded to consider more than one person making more than one decision. If the goal is to model how a group of people choose lunch over the course of a conference, there is the possibility of individual differences in the strategies that different people employ. Some people may care about every ingredient and item, while others focus on just one or two. There is also the possibility that, even if their underlying preferences remain the same, people will switch their strategies over time. The same person may on the first day, for example, consider all the components of the lunch box in detail, but on later days make decisions based on a less-detailed evaluation.

There is a large literature studying the strategies people use to make simple choices and developing and evaluating models of people's search and choice behavior (see Newell, Lagnado, & Shanks, 2015, chap. 3, for a summary). One general approach is to assume people tradeoff the effort involved in searching and integrating information with the increase in accuracy achieved (Beach & Mitchell, 1978; Fechner, Schooler, & Pachur, 2018; Lieder & Griffiths, 2017; Payne, Bettman, & Johnson, 1990). Another general approach is to assume that people rely on the structure of information in decision-making environments to limit the extent of their search and integration of information (Gigerenzer, Todd, & the ABC Group, 1999). Both of these approaches lead to models in which people may conduct limited search and may make noncompensatory decisions. These relatively simple strategies can then be contrasted with normative accounts involving exhaustive search and the rational combination of information.

The most thoroughly investigated strategies are those arising from the fast and frugal heuristics approach (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 1996; Gigerenzer et al., 1999). In particular, the strategy known as *take-the-best* has been tested extensively. According to take-the-best, people search features in order of decreasing cue validity—a measure of how often a cue indicates a correct decision—until a discriminating cue is found. The choice is then simply the one indicated by this first discriminating cue. Take-the-best is often contrasted with two more normative strategies that involve exhaustive search. In the *tally* strategy, people are assumed to count the number of features favoring each alternative and choose the one with the most features. In the *weighted-additive* strategy, people are assumed to combine all of the features for each alternative according to weights reflecting the validities of the features, so that more important features are given greater weight.

There is a large body of research that examines what sort of environments and task demands make these strategies more or less effective, and lead to them being used more or less often (Bergert & Nosofsky, 2007; Bobadilla-Suarez & Love, 2018; Bröder & Schiffer, 2006; Hilbig, 2008; Lee & Cummins, 2004; Lee, Newell, & Vandekerckhove, 2014; Mata, Schooler, & Rieskamp, 2007; Newell & Lee, 2011; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Rieskamp & Otto, 2006). This research has made progress in understanding the various factors that influence the strategies people use, although there remain many open questions. One clear regularity, however, is that there are large individual differences. It is not surprising, therefore, that the field has developed new methods for inferring individual differences, particularly by classifying people as using one strategy or another based on their overall choice behavior in a task. Many early studies relied on critical test trials that provided direct empirical measures of strategy use (Bröder, 2000; Lee & Cummins, 2004). More general methods for strategy inference were then developed. Bröder & Schiffer (2003a) developed an early Bayesian method for classifying strategy use. Lee (2004) developed a minimum description length approach using the normalized maximum likelihood (NML) criterion (Grünwald, 2007), which was later extended and applied by Newell and Lee (2011). Hilbig and Moshagen (2014) proposed a similar approach, also based on the NML. Most recently, the methodological emphasis has been on the use of latent-mixture models, in which two or more different strategies are included in a single general model as potential accounts of a person's decision making. Bayesian inference about the most likely mixture component, based on the observed decisions, corresponds to an inference about which strategy the person uses. The latent-mixture approach has emphasized the development of cognitive models of the data-generating processes, rather than treating the inference as strictly a data analysis problem (Lee, 2011, 2018). This approach was demonstrated by Lee (2016), building on the foundations laid out by Lee and Newell (2011; see also Lee & Wagenmakers, 2013, chapter 18; Scheibehenne, Rieskamp, & Wagenmakers, 2013) and further developed by Heck, Hilbig, and Moshagen (2017).

There has been much less empirical investigation of whether and how the same person changes strategies over a series of decisions. Most of the experimental manipulations that influence strategy use have been applied between-participants across different experimental conditions, although there are some exceptions (Bröder & Schiffer, 2006; Lee et al., 2014; Rieskamp & Otto, 2006). For example, Lee et al. (2014) manipulated the information structure of a cue environment so that it changed twice—from supporting noncompensatory to compensatory decision making, and then back again—without informing participants. They observed that most people changed strategies in response to these environmental changes, but that there were individual differences in the extent of these changes, consistent with different theoretical accounts of learning and adaptation. In contrast to the extensive methodological development for inferred differences between people in strategy use, there are no well-developed or widely used methods for inferring strategy shifts within the same individual.

One possible reason for the lack of exploration of strategy shifts is that it requires the ability to justify much more complicated accounts of people's behavior. More complicated models need more evidence, and most previous research has focused solely on the decisions people make for inferring strategy use. Where other sorts of behavior are examined, it is typically as an external test of inferences based on decision data. For example, Lee and Cummins (2004) classified participants as using either the take-the-best or the weighted-additive strategy based on their decisions, and then evaluated the extent to which confidence ratings and response times between these groups supported the original classification. Similarly, Walsh and Gluck (2016) classified participants as using either a compensatory or noncompensatory strategy based on their decisions, and then evaluated the extent to which search behavior and verbal reports were consistent with that classification. It is rare for additional sources of behavior beyond decisions to be used as the basis for making the classification in the first place, although there are a few exceptions. For example, Glöckner (2009, see also Jekel, Nicklisch, & Glöckner, 2010) developed a maximum-likelihood method for strategy inference that incorporated both decision and search information. Bergert and Nosofsky (2007) based inferences in a more specific application on both decisions and response times. Brandstätter, Gigerenzer, and Hertwig (2008) used search and decision to make inferences about strategy use, and Pachur, Hertwig, Gigerenzer, and Brandstätter (2013) and Lee et al. (2014) also used the extent of search as their primary behavioral measure in the specific models they considered. In general, though, most research in the literature relies exclusively on a set of binary decisions as the basis for inferring strategy use.

From a cognitive modeling perspective, the focus on one behavioral source of evidence is a limiting one. The goal of cognitive models is to describe, explain, and predict multiple sorts of behavior. Different strategies make different predictions about not just the decisions people will make, but also about the information search and cognitive processing they do before reaching those decisions. Thus, it should be possible to make more confident and detailed inferences about strategy use by using the available behavioral data. This sort of "common cause" modeling approach is widely used in the empirical sciences and provides a powerful way to make stronger inferences based on the available evidence (Lee, 2018).

In this article, we address both the challenge of using multiple behavioral data sources simultaneously to infer strategy use, and the challenge of inferring the number and timing of switches in strategy use within individuals. Using a previously reported data set (Walsh & Gluck, 2016), we extend the traditional approach of making inferences based on decisions alone to allow for multiple behavioral data sources. We show that this affects inferences about strategy use. We then develop a modeling approach that allows changes in strategy use over a sequence of trials to be inferred. The method infers how many strategy changes occur, at what trial they occur, and which strategies are used before and after the change. We apply this model to both the decision data and to the behavioral data, showing that there is evidence for all sorts of changes, especially when the behavioral evidence is used.

Behavioral Data

Our data come from an experiment reported by Walsh and Gluck (2016), involving a standard two-alternative forced-choice decision task in which participants explicitly searched for information (Bröder, 2000; Newell et al., 2003). In the Walsh and Gluck (2016) experiment, a total of 38 participants completed a sequence of 120 trials in which they chose between two alternative stocks. Participants had the option of considering information about the stocks prior to making their choice. That information was available in four binary attributes, each of which had a predictive validity associated with it. Both the order of attributes on the screen and the validity associated with each attribute were randomized across participants, so the specific stimulus details and their order were different for each participant.

On each trial, information about the stocks was limited to the four binary attributes, and participants could "buy" access to the information by clicking to reveal the covered attribute values. These mouse clicks provide data about information search processes. Walsh and Gluck (2016) included both "verbal report" and "silent" conditions as a between-subjects manipulation in their experiment. All of the participants completed verbalization training. Half of the participants were randomly assigned to a condition in which they provided concurrent verbal report data (Ericsson & Simon, 1993) as they searched and made decisions. This between-subjects manipulation allowed an evaluation of whether performance in the experiment was reactive to the requirement to provide verbal protocols.

Rigorous best practices were used to transcribe, segment, and code the verbal protocol data. The methodological details are previously published and available to the interested reader (Walsh & Gluck, 2016), and so are not repeated here. There are two key points relevant to those interested in the debate about the use of verbal protocols as a data source in this way. One point is that the coding system developed and used by Walsh and Gluck (2016) resulted in very high interrater reliability. The other point is that, consistent with the weight of evidence in the broader literature (Fox, Ericsson, & Best, 2011) the requirement to provide verbal protocols slowed response times slightly but did not appear to impact search or decision processes.

We use only the data from the 19 participants in the verbal report condition, because it provides three behavioral sources of evidence—decision, search, and verbal report—to understand the strategies participants use. Walsh and Gluck (2016) describe their study and the associated behavioral data, and we refer the interested reader to that article for more details.

Fixed Strategy: Inference From Decision Data

As discussed in the introductory section, the most common analysis for inferring people's use of heuristics in multiple-cue decision-making tasks is based solely on the decisions people make, and assumes they use the same strategy throughout the task. We implement this analysis first, as a reference point for the extended analyses that follow.

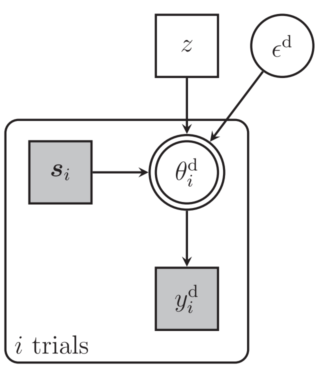
Graphical Model

A Bayesian graphical model for inferring strategy use, following the fixed-strategy assumption, is shown in Figure 1. Graphical models provide a formalism for expressing probabilistic generative models (Jordan, 2004; Pearl, 1998), and are well suited to the application of computational methods for Bayesian inference with cognitive models (Lee & Wagenmakers, 2013). In general, graphical models use nodes to represent parameters and data, and the graph structure indicates how the parameters generate the data.

In Figure 1 the data are the decisions made on each trial, represented by the y_i^{d} node for the ith trial, with the superscript d indicating the behavioral data are decisions. The node is square because the data are discrete: $y_i^{\mathrm{d}} = 1$ if the first alternative is chosen, and $y_i^{\mathrm{d}} = 0$ if the second alternative is chosen. The node is shaded because the data are observed: the values of every y_i^{d} are known. The cognitive model assumes these decision data are generated by the use of a strategy, z, which is applied imperfectly on each trial with some execution error ϵ^{d}. Besides the take-the-best, tally, and weighted-additive strategies that are our main focus, we also allow for a guessing strategy. This is naturally incorporated as an additional component of the latent-mixture model, and its inclusion provides a principled model-based approach to dealing with contaminant behavior (Zeigenfuse & Lee, 2010). The z node is square because there is a discrete number of alternatives: z can indicate the use of the take-the-best, tally, weighted-additive, or guessing strategy. The ϵ^{d} node is circular because it is continuous. Both the z and ϵ^{d} nodes are unshaded because they are latent or unobserved parameters: the main statistical goal of the graphical model is to make inferences about these psychological variables.

On each trial, the strategy being used and the execution error combine with the stimulus information for that trial—the cues for each alternative, and their validities, represented by the observed node $\mathbf{s}_i$—to produce a probability that the first alternative will be chosen. This trial-level process is placed within an encompassing plate in the graphical model. Plates represent independent replications of the graph structure and are used in Figure 1 to indicate the repetition of the same decision-making process on each trial.

The probability the first alternative is chosen on the ith trial is represented by θ_i^{d}. If a guess is being made, then each alternative has equal probability of being chosen. If the take-the-best, tally, or weighted-additive strategy is being used, then the probability of choosing the first alternative depends on the prediction made by that strategy. We represent this prediction as $\mathcal{H}_z^{\mathrm{d}}(\boldsymbol{s}_i)$ for strategy z, given the alternatives $\boldsymbol{s}_i$. Following the assumption of an execution error, this prediction is followed with (high) probability $(1 - \epsilon^{\mathrm{d}})$, but a random choice is made with execution-error probability ϵ^{d}. Thus, overall, the probability the first alternative will be chosen on the ith trial is



$$\begin{aligned}
z &\sim \text{Categorical}\left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right)\\
\epsilon^{\mathrm{d}} &\sim \text{Uniform}\left(0, \tfrac{1}{2}\right)\\
\theta_i^{\mathrm{d}} &= \begin{cases} \mathcal{H}_z^{\mathrm{d}}(\boldsymbol{s}_i)\left(1 - \epsilon^{\mathrm{d}}\right) + \frac{1}{2}\epsilon^{\mathrm{d}} & \text{if } z \in \{\text{ttb}, \text{tally}, \text{wadd}\}\\ \frac{1}{2} & \text{if } z = \text{guess} \end{cases}\\
y_i^{\mathrm{d}} &\sim \text{Bernoulli}\left(\theta_i^{\mathrm{d}}\right)
\end{aligned}$$

Figure 1. Graphical model for inferring the use of a single strategy across all trials based on decision data.

$$\theta_i^{\mathrm{d}} = \begin{cases} \mathcal{H}_z^{\mathrm{d}}(\boldsymbol{s}_i)\left(1-\epsilon^{\mathrm{d}}\right) + 1/2\epsilon^{\mathrm{d}} & \text{if } z \in \{\text{ttb, tally, wadd}\} \\ 1/2 & \text{if } z = \text{guess} \end{cases}$$

In the graphical model, the θ_i^{d} node is circular because it is a continuous probability and is double-bordered because it is deterministic. As the definition above makes clear, the choice probability is a function of other nodes, and is included for semantic clarity rather than out of statistical necessity. Given this probability for choosing the first alternative, the observed data are modeled as $y_i^{\mathrm{d}} \sim \text{Bernoulli}(\theta_i^{\mathrm{d}})$.

The coherence of Bayesian inference requires the specification of prior probabilities for latent parameters and provides a natural vehicle for formalizing key psychological assumptions of the model (Lee & Vanpaemel, 2018). For the execution error, we use the prior distribution $\epsilon^{\mathrm{d}} \sim \text{Uniform}\left(0, \frac{1}{2}\right)$, since this is the range of probabilities that give the error its intended meaning. For the strategies, we assume that $z \sim \text{Categorical}\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, so that they are all equally likely.

Individual Results

Figure 2 shows the trial-by-trial predictive accuracy of all four strategies for three selected participants, to provide concrete examples of the inferences from the graphical model.[1] Participant 1 is shown in the top panel, Participant 7 is shown in the middle panel, and Participant 3 is shown in the bottom panel. Within each panel, the four rows of markers correspond to take-the-best, tally, weighted-additive, and guessing strategies. The sequence of trials completed by the participants progresses from left to right. The area of each marker shows the predictive accuracy of a specific strategy for a specific trial. This area is maximal when the strategy predicts the binary choice made by the participant, is blank when it predicts the other choice, and is one half in those cases, for the tally and guessing strategies, where both choices are predicted to be equally likely.

The graphical model in Figure 1 infers that Participants 1 and 3 use the take-the-best strategy, but that Participant 7 uses the tally strategy. These inferences seem intuitively reasonable, based on the relative predictive accuracy of these strategies shown in Figure 2. For Participant 1, take-the-best correctly predicts decisions for every trial except Trial 51, and the other three strategies make much worse predictions. For Participant 7, the tally strategy only mispredicts for five trials—Trials 1, 84, 87, 94, and 103—although there are many trials for which both alternatives have an equal number of cues, and so the tally strategy predicts that both choices are equally likely. Nonetheless, the other three strategies are clearly less accurate in their predictions, and the inference that Participant 7 uses the tally strategy makes sense. For Participant 3, take-the-best fails to predict eight of the first 24 trials, but then correctly predicts the choices made on the remaining trials.

Overall Results

The results of applying the graphical model in Figure 1 to the participants are shown in Figure 3. Each bar corresponds to a participant, and the composition of each bar corresponds to the posterior probability that each of the four possible strategies was the one that generated that participant's decision data. For 11 of the participants the take-the-best strategy is inferred with complete certainty. For Participant 7 the tally strategy is inferred with complete certainty. For the remaining seven participants there is uncertainty. In four of these cases, the uncertainty is between the tally and weighted-additive strategies, both of which are compensatory strategies.

It is important to understand the distributions in Figure 3 represent uncertainty in which strategy is used, and not variability in strategy use. The assumption of the model is that each participant uses one strategy throughout the experiment. The inference for Participant 19, for example, is that most likely this was the take-the-best strategy, but there is also evidence consistent with the use of the tally strategy. The inference is not that the participant uses some combination of these strategies to make their decisions. Rather, given the evidence available in the decisions made across all 120 trials, there is a high probability that Participant 19 uses the

[1] Note that these are genuine predictions, because they are formed without reference to the behavioral data. Cognitive modeling has an unfortunate habit as a field of describing "fitted" data as "model predictions," in violation of the foundations of empirical science (Feynman, 1994). For a discussion, see Lee (2018), especially the section "Fitting data is not the same as predicting data," and Roberts and Pashler (2000, 2002).

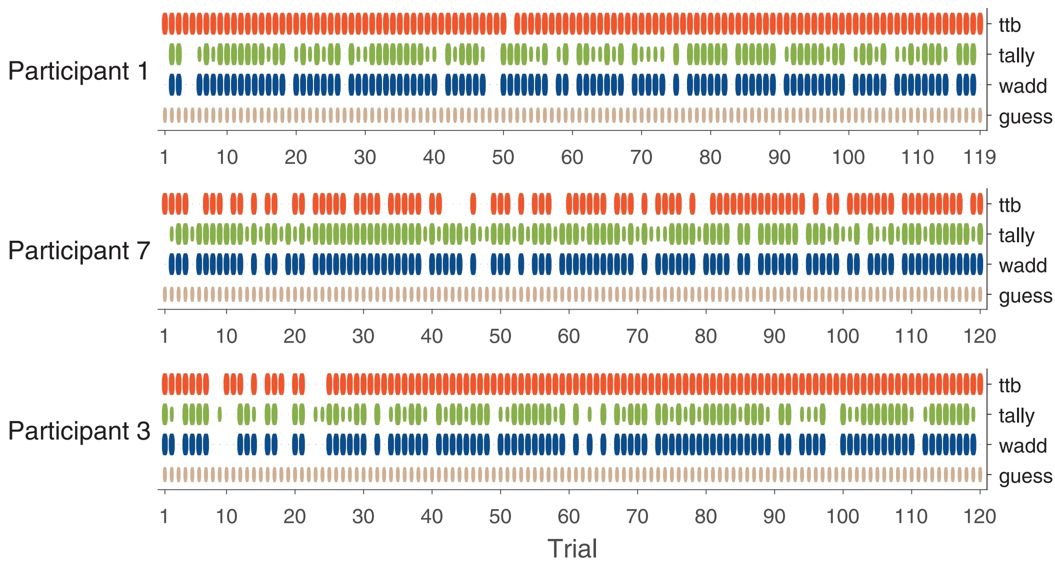


Figure 2. Predictive accuracy for the take-the-best, tally, weighted-additive, and guessing strategies on each trial for Participants 1, 7, and 3. The area of the markers show the predictive accuracy of the strategy for that trial. See the online article for the color version of this figure.

take-the-best strategy, a low probability that they use the tally strategy, and no evidence that they use either the weighted-additive or guessing strategies.

Fixed Strategy: Inference From All Behavioral Data

In the preceding analysis, we limited the behavioral data strictly to the decision data. That is, to make inferences about which strategy each participant used, we considered only which stock was picked, and not what cues were searched or what verbal reports were made. As we noted earlier, this is consistent with standard practice. The first major goal of this article is to overcome this limitation, and base inferences about strategy use on multiple sources of behavioral evidence. The generative modeling approach makes this extension conceptually clear. The idea is that the use of each strategy makes predictions about each type of behavior, and so each source of behavioral evidence can inform inferences about strategy use. The simplest assumption is that the strategy a participant uses is a common cause for their search behavior and the verbal statements they make, as well as for their decisions. An advantage of the Bayesian graphical modeling approach is that the model in Figure 1 can easily be extended hierarchically

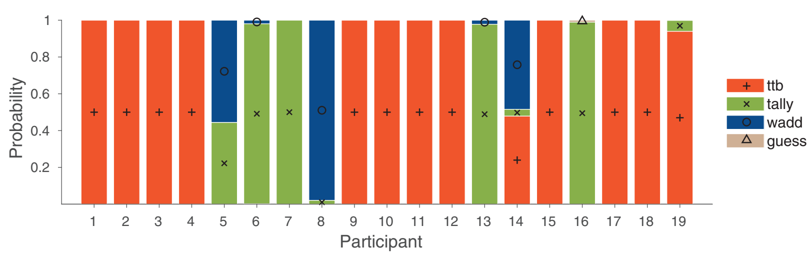


Figure 3. Inferences about fixed-strategy use for all 19 participants based on decision data. See the online article for the color version of this figure.

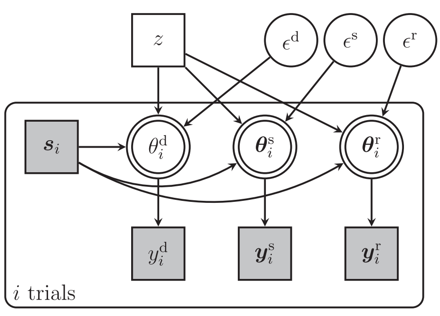
to implement the common-cause assumption (Lee, 2018).

Graphical Model

Figure 4 shows an extended graphical model that assumes the same strategy generates decisions, search behavior, and verbal reports. On each trial, there is now a decision y_i^{d}, a set of search behaviors $\mathbf{y}_i^{\text{s}}$, and a set of verbal reports $\mathbf{y}_i^{\text{r}}$. For search behavior, $y_{ik}^{\text{s}} = 1$ if the kth cue was searched on the ith trial, and $y_{ik}^{\text{s}} = 0$ if it was not. For our data, k ranges from 1 to 8, covering the four cues for the two alternatives, each of which could be searched independently. For report behavior, $y_{ik}^{\text{r}} = 1$ if the kth type of verbal report was made on the ith trial, and $y_{ik}^{\text{r}} = 0$ if it was not. For our data, k ranges from 1 to 15, covering the different verbal response categories used by Walsh and Gluck (2016).

The graphical model in Figure 4 assumes that each behavioral data type has its own execution error given, respectively, by ϵ^{d}, ϵ^{s}, and ϵ^{r}. As for the decision data, the search and report execution errors control the probabilistic application of the predictions of the strategy z to each trial. The search predictions $\mathcal{H}_z^{\text{d}}(\boldsymbol{s}_i)$ for strategy z, given the alternatives $\mathbf{s}_i$, apply to all eight possible search behaviors. For the compensatory tally and weighted-additive strategies, the predictions are that every cue will be searched and that search order is irrelevant. For the take-the-best strategy, the prediction is that cues will be searched in order of decreasing cue validity until a discriminating cue is found, at which point search terminates. These predictions are followed with an execution error, ϵ^{s}, so that a specific cue for a specific alternative is searched with probability $(1 - \epsilon^{\text{s}})$ if it is predicted to be searched, and with probability ϵ^{s} if the prediction is that it will not be searched. For the guessing strategy, we assume that each of the eight possible search actions has probability $\frac{1}{4}$. This is intended to formalize the assumption that any cue for any alternative is equally likely to be searched but that the overall extent of search will be limited. The probability of $\frac{1}{4}$ corresponds, on average, to searching one cue from each alternative, as for the most frugal reasonable search strategy.

It is important to note that we model each possible search behavior, in the form of choosing to reveal a specific cue for a specific alternative, separately. The different strategies make predictions about whether each of these search behaviors will occur, and our generative model assumes that each specific behavior is followed according to the execution error ϵ^{s}. This means that our model does not require that the entire pattern of search behavior predicted by a strategy be followed in order to find evidence in favor of the use of that strategy. For example, it is not necessary for every cue for both alternatives to be searched in order to find evidence for the tally or weighted-additive strategies. Rather, the more extensive the search behavior is, the more evidence will be accrued for these strate-



$$
\begin{aligned}
z &\sim \text{Categorical}\left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right) \\
\epsilon^{\text{d}}, \epsilon^{\text{s}}, \epsilon^{\text{r}} &\sim \text{Uniform}\left(0, \tfrac{1}{2}\right) \\
\theta_i^{\text{d}} &= \begin{cases} \mathcal{H}_z^{\text{d}}(\boldsymbol{s}_i)\left(1-\epsilon^{\text{d}}\right) + \tfrac{1}{2}\epsilon^{\text{d}} & \text{if } z \in \{\text{ttb}, \text{tally}, \text{wadd}\} \\ \tfrac{1}{2} & \text{if } z = \text{guess} \end{cases} \\
\theta_{ik}^{\text{s}} &= \begin{cases} (1-\epsilon^{\text{s}}) & \text{if } z \in \{\text{ttb}, \text{tally}, \text{wadd}\}, \mathcal{H}_z^{\text{s}}(\boldsymbol{s}_i) = 1 \\ \epsilon^{\text{s}} & \text{if } z \in \{\text{ttb}, \text{tally}, \text{wadd}\}, \mathcal{H}_z^{\text{s}}(\boldsymbol{s}_i) = 0 \\ \tfrac{1}{4} & \text{if } z = \text{guess} \end{cases} \\
\theta_{ik}^{\text{r}} &= \begin{cases} (1-\epsilon^{\text{r}}) & \text{if } \mathcal{H}_z^{\text{r}}(\boldsymbol{s}_i) = 1 \\ \epsilon^{\text{r}} & \text{if } \mathcal{H}_z^{\text{r}}(\boldsymbol{s}_i) = 0 \\ \tfrac{1}{2} & \text{if } \mathcal{H}_z^{\text{r}}(\boldsymbol{s}_i) = \tfrac{1}{2} \end{cases} \\
y_i^{\text{d}} &\sim \text{Bernoulli}\left(\theta_i^{\text{d}}\right) \\
y_{ik}^{\text{s}} &\sim \text{Bernoulli}\left(\theta_{ik}^{\text{s}}\right) \\
y_{ik}^{\text{r}} &\sim \text{Bernoulli}\left(\theta_{ik}^{\text{r}}\right)
\end{aligned}
$$

Figure 4. Graphical model for inferring the use of a single strategy across all trials, based on decision, search, and verbal report data.

gies. In effect, modeling search in terms of individual cues, rather than complete patterns of search, allows for search patterns that are merely similar to those strictly predicted by the strategies to correspond to partial evidence in favor of the strategies.

To relate the strategies to the verbal report data, we make assumptions about which verbal report categories each strategy predicts will be produced. These modeling assumptions are detailed in Table 1. For many combinations of strategy and report type, there is no obvious prediction, indicated by a "–" symbol. For some combinations, however, there is a clear prediction. For example, a verbal report of the "decision: single indicator" type is predicted by take-the-best, but not by the compensatory tally and weighted-additive strategies. Those combinations where verbal reports are predicted to be produced are indicated by a 1, and those combinations where the verbal report is predicted not to be produced are indicated by 0.

It is clear from Table 1 that only six of the verbal report types provide evidence for inferring the use of one strategy rather than another. The "elaboration: indicator based" and "decision: single indicator" favor take-the-best, the "elaboration: stock based" and "elaboration: holistic" favor the compensatory strategies, and "decision: guess" is consistent with the guessing strategy. Table 2 provides example statements for each of these six discriminating verbalization types.

To use the evidence provided by the verbal report data, we make a slightly different assumption than for the decision and search data. In particular, we model only those verbal reports that are observed. This means that, unlike the decision and search data, we do not treat the absence of a report predicted by a strategy as evidence against the use of that strategy. We think this is a reasonable approach. Although participants are required to make a decision and to search as extensively as they needed to make that decision, there is not the same compulsion to choose whether or not to produce every possible sort of verbal report on every trial (Ericsson & Simon, 1993). It may be that participants had thoughts consistent with verbal report types that they sometimes did not produce, and so there is little evidence in the absence of a report. Instead, we focus on the evidence provided by the reports actually produced. Formally, the produced reports $\mathbf{y}_i^r$ have probability $(1 - \epsilon^r)$ if they are predicted by strategy z, probability ϵ^r if they are not predicted by strategy z, and probability $\frac{1}{2}$ if there is no prediction.

We use the priors $\epsilon^d \sim \text{Uniform}\left(0, \frac{1}{2}\right)$, $\epsilon^s \sim \text{Uniform}\left(0, \frac{1}{2}\right)$, and $\epsilon^r \sim \text{Uniform}\left(0, \frac{1}{2}\right)$ for the three execution error parameters. The statistical assumption of different execution accuracies for

Table 1
Model Probabilities of the Take-the-Best, Tally, Weighted-Additive, and Guessing Strategies Generating the 15 Report Types Used by Walsh and Gluck (2016)

Report type	Take-the-best	Tally	Weighted-additive	Guess
Search				
Encoding				
Evaluation				
Metacognitive				
Miscellaneous				
Uncategorized				
Elaboration: Indicator-based	1	0	0	
Elaboration: Stock-based	0	1	1	
Elaboration: Holistic	0	1	1	
Elaboration: Information value				
Decision: No justification				
Decision: Single indicator	1	0	0	
Decision: Multi-indicator	0	1	1	
Decision: Memory				
Decision: Guess				1

Table 2
Example Statements for Each Type of Verbalization Associated With Compensatory and Noncompensatory Decision Strategies

Report type	Example statement
	Noncompensatory
Elaboration: Indicator-based	"Financial reserves for one is better."
	"Both are yes for financial reserves."
Decision: Single indicator	"Pick the one who has reserves."
	"I'm going with TJM because of the share trend."
	Compensatory
Elaboration: Stock-based	"One of them is both established and the share trend has been going up."
	"DTJ does not have a yes."
Elaboration: Holistic	"So we're split between the two of them."
	"They're identical."
Decision: Multi-indicator	"I'll go with this one since it is yes for both of them."
	"So they are positive and established."
	Guess
Decision: Guess	"I guess I'll just guess."
	"So I'll take a 50/50."

Note. This is a simulated stock selection task. TJM and DTJ are imaginary stock ticker symbols used as stimuli in the task to represent different stock options.

each source of behavioral data corresponds to the psychological assumption that decisions, search, and verbal reports may have different reliabilities or levels of usefulness in terms of how closely they adhere to the deterministic predictions of the underlying strategies. In terms of inference, this means that the different behaviors can have different levels of importance in determining the inferences made about strategy use. Meanwhile, the statistical assumption of the execution errors being a priori independent corresponds to the psychological assumption that any pattern of reliability or usefulness across the behaviors is possible. The Bayesian modeling framework could, however, accommodate stronger theoretical assumptions. For example, the theoretical assumption that decision data are more informative than verbal report data could be formalized as a constraint on the joint prior distributions of ϵ^d and ϵ^r.

Individual Results

Figure 5 shows the decision, search, and report behavior for Participant 6, to provide a concrete example of how examining multiple behavioral sources can affect inference about strategy use. The bottom panel shows the predictive accuracy of the four strategies for each decision. Figure 3 shows that the inference based on just the decisions made by Participant 6 is that the tally strategy is being used. This seems intuitively reasonable, on the grounds that—although none of the strategies predict the sequence of decisions made by Participant 6 very well—the tally strategy only mispredicts 25 of those decisions, compared with 41 and 39 mispredictions for the take-the-best and weighted-additive strategies.

The middle panel of Figure 5 summarizes the search behavior for the participant. For each trial, the height of the bar shows the number of cues searched on that trial, out of the maximum possible of eight. A black dot indicates the point in the search process at which the first discriminating cue was found. Thus, bars with dots at their peak are consistent with take-the-best search, while bars that extend beyond the dot are consistent with more extensive search under the tally and weighted-additive strategies. What is seen in Figure 5, however, is that the participant often did not even conduct an extensive enough search to find a single discriminating cue. On almost all trials, exactly two cues are searched, and only a minority of trials involve the discovery of a discriminating cue. This search behavior is more consistent with the guessing strategy.

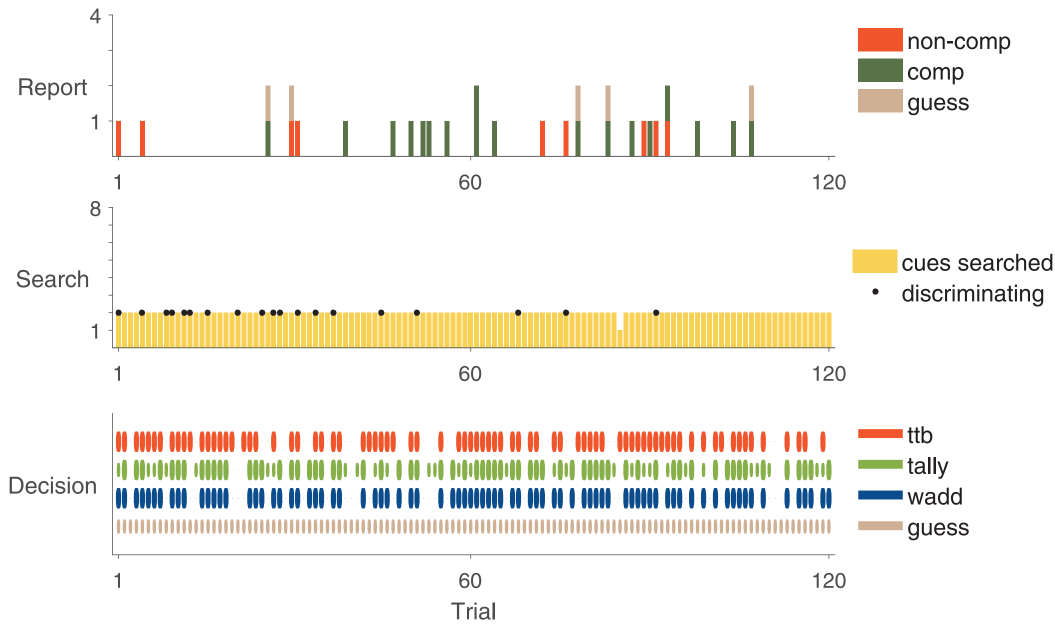


Figure 5. The decision, search, and verbal report behavior of Participant 6. The bottom panel shows the predictive accuracy for the take-the-best, tally, weighted-additive, and guessing strategies on each trial. The bars in the middle panel show the number of cues searched on each trial, and the filled circles indicate finding a discriminating cue. The bars in the top panel show the number of verbal reports consistent with compensatory and noncompensatory strategies for each trial. See the online article for the color version of this figure.

The top panel of Figure 5 summarizes the verbal report behavior for Participant 6, based on the assumed relationships between strategies and report types detailed in Table 1. Because the tally and weighted-additive strategies are assumed to relate to the report types in the same way, they are treated as a single "compensatory" class. Take-the-best is correspondingly labeled as "non-compensatory," and the guessing strategy maintains its own class. The stacked bars count the number of verbal reports belonging to each class for each trial. There are never more than two of these reports on a single trial, and they are distributed over the three possible classes.

Based on the additional behavioral evidence provided by the search and report data, the graphical model in Figure 4 infers that Participant 6 is almost certainly using the guessing strategy. This reversal of the original tally strategy inference, based on the decision data, makes sense. The additional evidence provided by the search and report behavior is not well predicted by the tally strategy, but it is more consistent with guessing. In particular, the prediction of the tally strategy is that every cue will be searched on every trial, which is clearly inconsistent with the observed behavior of the participant. Instead, the search behavior is highly consistent with the predictions of the guessing strategy, which assumes a small number of cues will be searched, without regard to whether they reveal a discriminating cue.

Participant 6 provides a good example of how additional behavioral evidence can lead to different inferences about strategy use. The question of which inference is "correct" is a subtle one. Statistically, the correct inference is the one that follows from the available information and modeling assumptions. From this perspective, both the tally strategy inference based on the decision data, and the guessing strategy inference based on all the behavioral data, are "correct." They both follow from the modeling assumptions made explicit in their respective graphical models, and the relevant set of behaviors produced by the participant. Psychologically, we argue that the correct inference is the one based on more complete behavioral evidence, at least given an acceptance of the common-cause assumption that the same strategy underlies the decision, search, and verbal report behaviors. Models of psychological phenomena should be as complete as possible, and the ability of the guessing strategy to account for all three sources of behavioral evidence—especially its significantly superior account of search behavior—makes it a better inference for the strategy used by Participant 6.

Overall Results

The result of applying this model to all of the participants is shown in Figure 6. Once again, each bar corresponds to a participant, and the composition of each bar corresponds to the posterior probability that each of the four possible strategies was the one that generated that participant's decision data. There are now 11 participants inferred with certainty to use take-the-best, one using tally, two using weighted-additive, and three using the guessing strategy.

For the majority of participants, the inferences are consistent with those based on only decision data, as shown in Figure 3. The notable exceptions are Participant 6, as already discussed; Participants 11 and 16, who are inferred to be using a guessing strategy rather than a substantive strategy; Participant 12, who is inferred to use the weighted-additive rather than the take-the-best strategy; and Participant 13, who is inferred to use take-the-best rather than the tally strategy.

Switching Strategies: Inference From Decision Data

The second major goal of this paper is to allow for switches in the strategy the same individual uses over a sequence of trials. We now tackle this goal, returning first to the case in which strategy inferences are based solely on decision data. Changing strategy use could take many forms, but perhaps the simplest starting assumption is one of discrete change. This would mean that people use a strategy for a sequence of trials, but, at some point, switch to a different strategy. It would be possible to make theoretical assumptions about how often these switches occur, and the likelihood of different transitions between strategies, such as from compensatory to noncompensatory strategies. We restrict ourselves, however, to the most basic model that simply allows for some unknown number of switches and does not constrain which strategies are used or what sort of transitions between strategies are common.

Graphical Model

The graphical model in Figure 7 uses only decision data but incorporates the possibility of change over the course of the trials. This involves the introduction of new parameters $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_\gamma)$ representing the trials at which the strategy changes, and the extension of the strategy parameter z to $\mathbf{z} = (z_1, \ldots, z_\gamma)$ representing the strategy used between the change points. The switch points are ordered, so that strategy z_1 is used from trial 1 to trial τ_1, strategy z_2 is used from trial $\tau_1 + 1$ to trial τ_2, and so on. The new deterministic variable w_i corresponds to the strategy being used on the ith trial.

The number of possible change points γ is set to be larger than is needed, and a key feature of the model in Figure 7 is that it infers the appropriate number of change points. This is done using an extension of a method developed and demonstrated by Lee (in press), for which more details are given in

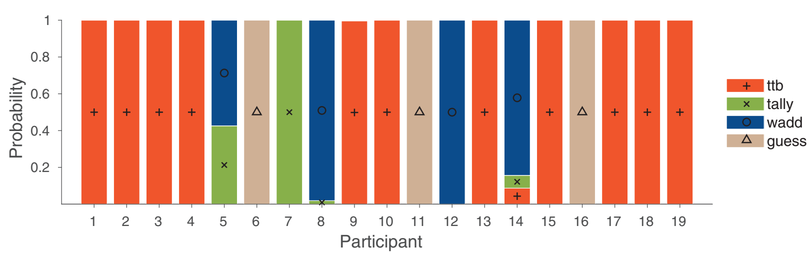
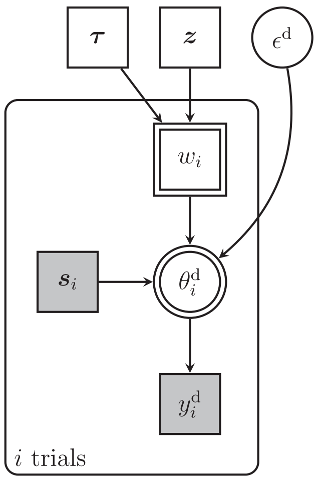


Figure 6. Inferences about fixed-strategy use for all 19 participants based on decision, search, and verbal report data. See the online article for the color version of this figure.



$$
\begin{aligned}
\tau'_k &\sim \text{Categorical}\Big(\overbrace{\frac{n}{2n-1}, \frac{1}{2n-1}, \ldots, \frac{1}{2n-1}}^{n \text{ trials}}\Big)\\
\boldsymbol{\tau} &= \text{sort}\big(\boldsymbol{\tau}'\big)\\
z_k &\sim \text{Categorical}\big(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\big)\\
\epsilon^{\mathrm{d}} &\sim \text{Uniform}\big(0, \tfrac{1}{2}\big)\\
w_i &= z_x, \text{ where } x = \textstyle\sum_k \mathcal{I}\left(i \geq \tau_k\right)\\
\theta_i^{\mathrm{d}} &= \begin{cases} \mathcal{H}^{\mathrm{d}}_{w_i}\left(\boldsymbol{s}_i\right)\left(1-\epsilon^{\mathrm{d}}\right) + \frac{1}{2}\epsilon^{\mathrm{d}} & \text{if } w_i \in \{\text{ttb}, \text{tally}, \text{wadd}\}\\ \frac{1}{2} & \text{if } w_i = \text{guess} \end{cases}\\
y_i^{\mathrm{d}} &\sim \text{Bernoulli}\big(\theta_i^{\mathrm{d}}\big)
\end{aligned}
$$

Figure 7. Graphical model for inferring strategy switches across trials on the basis of decision data.

Appendix A. The key to the method is the prior $\tau'_k \sim \text{Categorical}\big(\frac{n}{2n-1}, \frac{1}{2n-1}, \ldots, \frac{1}{2n-1}\big)$, where n is the number of trials. The $\boldsymbol{\tau}'$ parameters are the nonorder-constrained switch points. By design, the prior on τ'_k gives half the prior mass to a change point at the first trial, which does not impact the predictions made by the model. Only change points from the second trial onward affect predictions. For these change points at Trials 2 to n, the prior mass is set to be equal.

In this way, the prior leads to posterior inference accomplishing both model selection and parameter estimation goals. The model selection goal is to decide whether a change point exists. This is accomplished by examining whether the change point is inferred to be at Trial 1 (i.e., the "does not exist" model), or at any other trial (i.e., the "does exist" model), and both models are given equal prior probability. The subsequent parameter estimation goal is, if the change point does exist, to determine at what trial it occurs. This is accomplished by the posterior probability over the Trials 2 to n, all of which are equally likely a priori.

Individual Results

Figure 8 summarizes the results of inferred strategy shifts for Participants 1, 7, and 3. These are the same three participants considered in Figure 2, under the previous assumption that the same fixed strategy is used on every trial. Figure 8 shows the same information about the predictive accuracy of each strategy on each trial. The new inference about potential switches in strategy use is shown by the bar immediately above the decision predictions.

For Participant 1, the inference remains unchanged: the take-the-best strategy is used for all trials. For Participant 7, however, two switch points are inferred, at Trials 84 and 95. Between these trials, the participant is inferred to use the take-the-best strategy. Before and after these switch points, however, they are inferred to use the tally strategy. The predominant use of the tally strategy is consistent with the inference made under the assumption of a single strategy and makes sense in terms of the predictive accuracies shown. The inferred block of using the take-the-best strategy also, however, makes sense, based on these predictive accuracies. Between Trials 84 and 95, the take-the-best strategy perfectly predicts the choices of the participants, and two of the four mispredictions of the tally strategy occur.

Two switch points are also inferred for Participant 3. For the first seven trials, the use of the take-the-best strategy is inferred, before a sequence of guessing trials, and a return to the take-the-best strategy after 25 trials. These in-

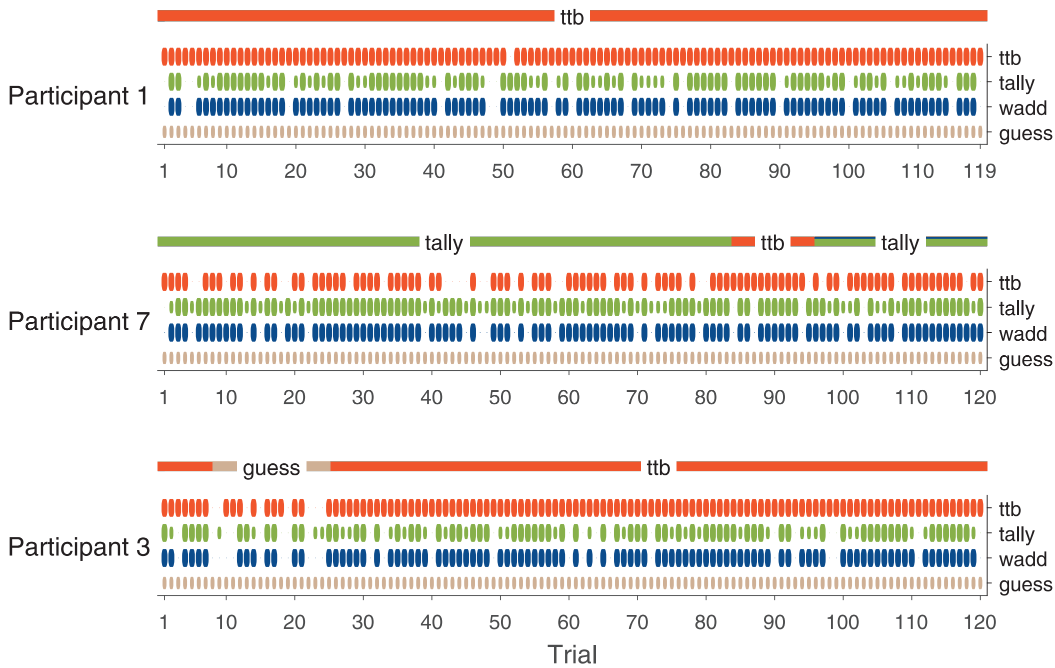


Figure 8. Inferences about strategy switching for Participants 1, 7, and 3. Each panel corresponds to a participant, and the predictive accuracy of each strategy on each trial is shown. The horizontal bar immediately above shows the inferred strategy, and changes in strategy, for the sequence of trials. See the online article for the color version of this figure.

ferences about strategy switching also seem reasonable, given the predictive accuracies of the strategies. The take-the-best strategy accounts perfectly for the choices of the participant for most of the experiment, except for the one early block of trials for which none of the substantive strategies provide accurate predictions.

Overall Results

Figure 9 shows the inferences made about switching strategy use for the participants, based on the decision data. Each horizontal bar corresponds to a participant and extends from left to right from the first to last trial of the decision-making task. The horizontal components of the bar correspond to strategy use between inferred switch points. The vertical components within these blocks correspond to uncertainty about which strategy was used. For example, Participant 14 is inferred to switch strategy once, around Trial 80. Up until that trial, it is uncertain whether they use a tally or weighted-additive strategy, although it is more likely they use the tally strategy. After the change point around Trial 80, the inference is that they use take-the-best for the remainder of the trials. A similar uncertainty in early trials is shown in Figure 9 for Participants 5 and 9, although the mix and weight of inferred strategies is different. Sometimes the strategy uncertainty occurs later in the experiment, as is the case for Participants 7 and 12.

Figure 9 shows that at least one switch point is inferred for eight out of the 19 participants. The strength of evidence provided by the data for the existence of these switch points can be quantified by Bayes factors (BF; Wagenmakers, Morey, & Lee, 2016), comparing the "null" fixed-strategy model formalized in Figure 1 with the general model that allows for strategy switches formalized in Figure 7. These BFs are shown for each participant in Figure 9, expressed as evidence in favor of the preferred model. Thus, for example, the BF = 1/8 result for Participant 1 means that their decision data provide eight times as much evidence for the

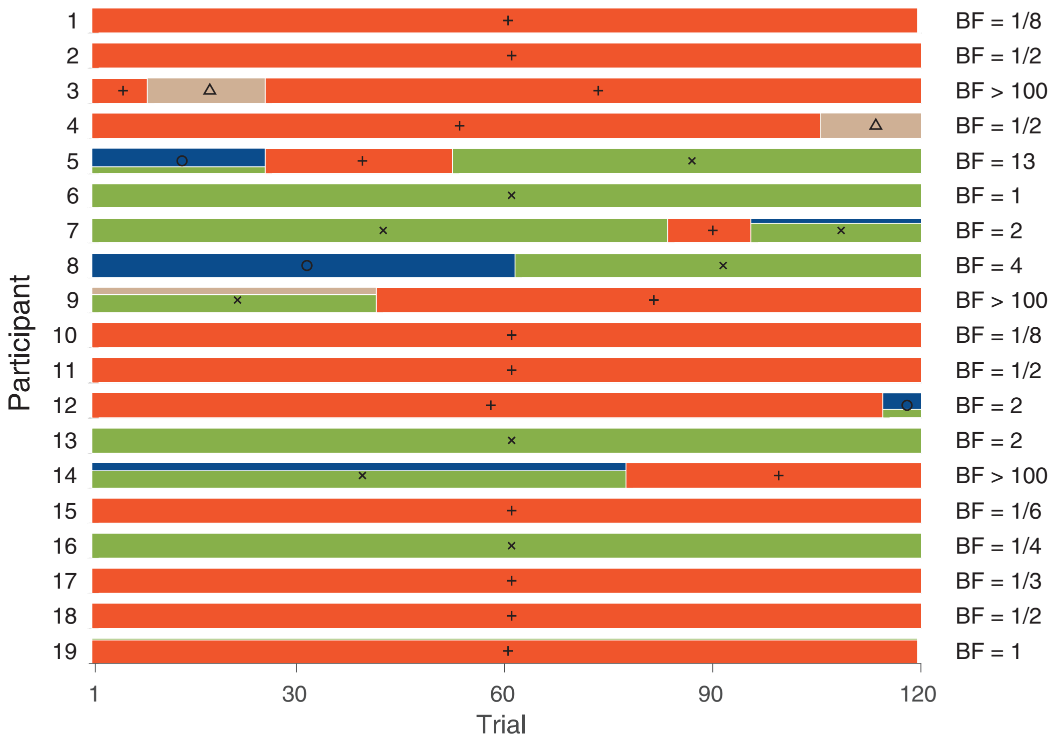


Figure 9. Question: Inferences about switching strategy use for all 19 participants based on decision data. Red = take-the-best (+ marker); green = tally (× marker); blue = weighted additive (circle marker); beige = guess (triangle marker). See the online article for the color version of this figure.

use of a fixed strategy, compared to the use of multiple strategies with switch points. For Participant 8, however, the BF = 4 means that their decision data provide four times as much evidence for a switch in strategy, compared to the use of a single fixed strategy. For many of the participants, the BFs do not provide overwhelming evidence either for or against switching strategies. Participants 3, 9, and 14 are exceptions, with BFs over 100 in favor of there being switches in their strategies.

For the participants for whom no switching was inferred, the inference about their strategy reduces to those made by the original fixed strategy model, as presented in Figure 3. The eight participants inferred to switch strategies show a range of different patterns. Several participants have early or late blocks of guessing trials, which could be associated with task familiarization and task boredom, respectively. In a number of cases, the transition from one strategy to another is from a cognitively more complicated strategy to a cognitively less complicated one, such as switching from the weighted-additive to the tally strategy or switching from the tally strategy to take-the-best.

Prediction and Generalization Performance

As a final evaluation of the model that allows for switching strategies, we consider its prediction and generalization performance (Busemeyer & Wang, 2000). Both prediction and generalization provide external tests of models, through their use of unobserved data. This is important both as a check on the model selection results, including the BFs already presented, and as a way of gauging the absolute, rather than relative, performance of the models.

By prediction, we mean the ability of a model to predict unseen behavior of the same type that has been observed. To do this, we test the ability of the fixed and switching strategy models to predict the decision made by a participant on the next trial based on the decisions made by the participant on all preceding trials. By generalization, we mean the ability of the model to predict unseen behavior of a different type. To do this, we test the ability of the fixed and switching strategy models to predict the search behavior made by a participant on the current trial, based on the decisions made by the participant up to that trial, but never having observed any search data. We do not consider generalization to the verbal reports, because we have not specified a complete generative model for these data. Without a model of when verbal reports are produced, a generalization test is not possible.

Figure 10 shows the results of these prediction and generalization tests, for the eight participants who were inferred to switch strategies based on their decisions. The top panels correspond to the decision prediction test, with each panel corresponding to the individual participant indicated. Within a panel, the smoothed log-likelihood prediction of the next decision is shown by the line with circular markers for the strategy switching model, and by the solid line for the fixed strategy model. Here the predictions are smoothed by averaging them over a window that extends 20 trials before and after the current trial. The coloring of the circular markers and line shows the inferred strategy over the trials for the switching strategy model. Broken vertical lines show the trials at which a strategy shift was inferred. The color of the solid line shows the inferred strategy for the fixed strategy model. A solid line bounded by square markers shows the log-likelihood predictive accuracy of a chance model that assumes each decision is equally likely on each trial.

It is clear from the decision prediction panels on Figure 10 that the strategy switching model almost always makes predictions that are as accurate, and often more accurate, than the fixed strategy model. It is also clear that both models generally perform much better than chance. Sometimes this superiority in prediction stems from the ability of the strategy switching model to identify when a different strategy is used. A good example of this is provided by Participant 14, since the increase in predictive accuracy is largest once the switch from the tally to the take-the-best strategy is inferred near Trial 80. More generally, however, the strategy switching model benefits from being able to make more confident predictions because it infers a smaller execution error. This is why, for example, the strategy switching model has a better log-likelihood than the fixed strategy model for Participants 3 and 4, even when both models are predicting a decision consistent with the take-the-best strategy for a long sequence of trials. The switching strategy model accounts better for early trials for Participant 3, and late trials for Participant 4, which means a lower execution error is inferred. This lower execution error, in turn, increases the agreement between the strategy switching model and the sequence of decisions consistent with the take-the-best strategy.

The bottom panels of Figure 10 correspond to the search generalization test, with each panel again corresponding to an individual participant. The smoothed log-likelihood prediction of the search behavior on the current trial is shown by the line with circular markers for the strategy switching model, and by the solid line for the fixed strategy model. The coloring of these lines indicate the inferred strategies for each model. Because the models never observe search data, their inferences about the execution error involved in search are only informed by prior assumption, which is vague. Accordingly, we do not attempt to quantify the chance model as an absolute benchmark for the search generalization test but focus on comparing the two models. Figure 10 shows that the differences in generalization accuracy are relatively small, both in magnitude and frequency. The two largest differences in generalization to search behavior, for Participants 3 and 14, both favor the strategy switching model for sequences of trials in which strategy switches were inferred.

Switching Strategies: Inference From All Behavioral Data

The two extensions of the standard analysis we have considered involve using multiple sources of behavioral data and allowing for switching strategies. These two extensions are independent of one another and so are simple to implement simultaneously. Accordingly, our final approach combines the two goals of the paper, providing a model that infers the number and type of strategy switches people make, based simultaneously on their decision, search, and verbal report behavior.

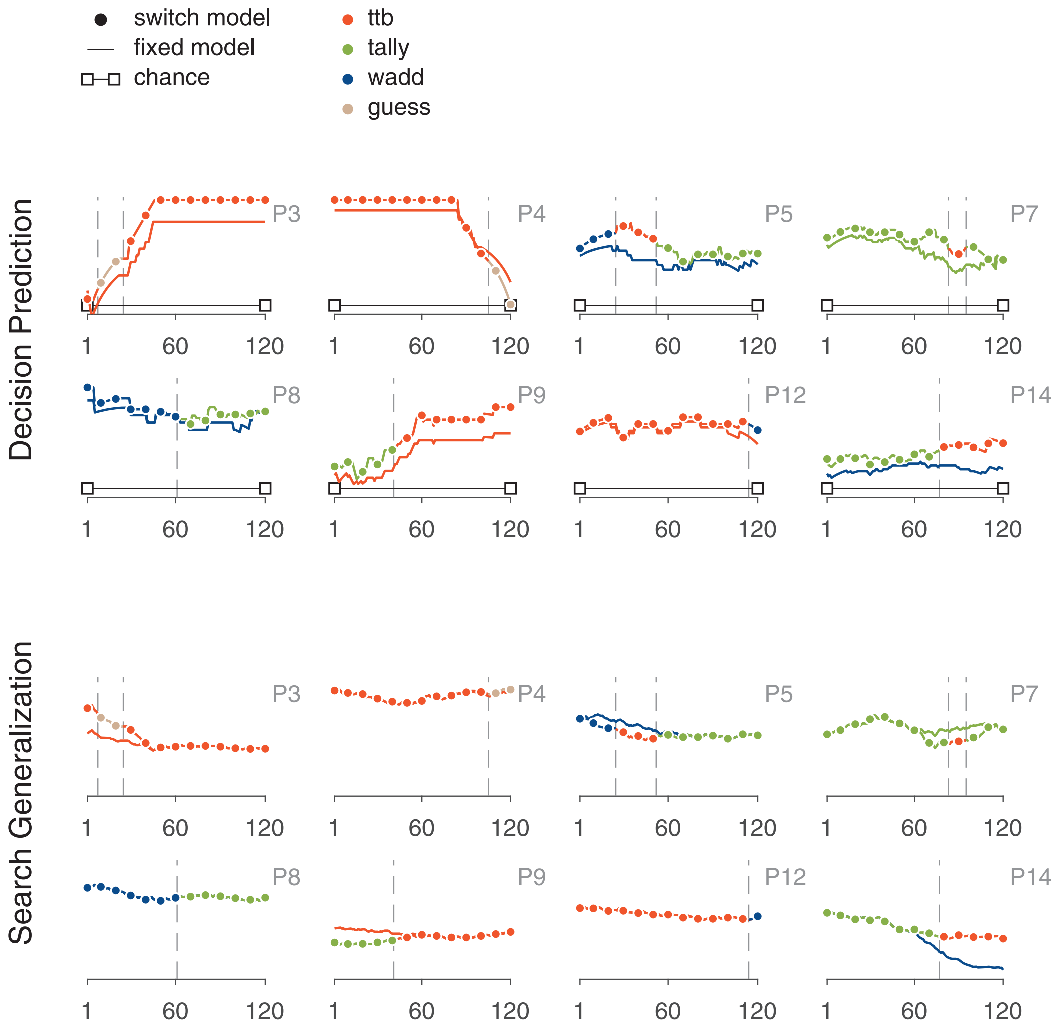


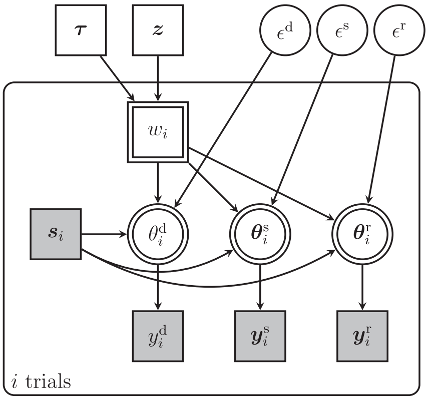
Figure 10. Predictive performance for decisions and generalization performance for search behavior for the model assuming switching strategies, the model assuming fixed strategies, and a chance model, based on decision data. The top panels relate to the ability of the models to predict the decision on the next trial, based on observing all behavior up to the preceding trial. The bottom panels relate to the ability of the models to generalize to the unobserved search behavior on the current trial, based on observing all decisions up to the current trial. Individual panels correspond to participants, and lines show the smoothed predictive performance over the experimental trials. See the online article for the color version of this figure.

Graphical Model

The graphical model for this final approach is shown in Figure 11. It combines the graphical models in Figure 4 and Figure 7. The switch points $\boldsymbol{\tau}$ and strategies $\boldsymbol{z}$ combine to determine w_i, the strategy being used on the ith trial, as in Figure 7. This strategy then generates behavior on the decision, search, and verbal report data for the ith trial, incorporating the execution errors, as in Figure 4.

Individual Results

Figure 12 summarizes the results of applying this model to Participant 8. The decision, search, and report panels show their observed behavior over the trials. The bars above these



$$
\begin{aligned}
\tau_k &\sim \text{Categorical}\Big(\overbrace{\frac{n}{2n-1}, \frac{1}{2n-1}, \ldots, \frac{1}{2n-1}}^{n \text{ trials}}\Big) \\
z_k &\sim \text{Categorical}\left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right) \\
\boldsymbol{\tau} &= \text{sort}\left(\boldsymbol{\tau}'\right) \\
\epsilon^{\mathrm{d}}, \epsilon^{\mathrm{s}}, \epsilon^{\mathrm{r}} &\sim \text{Uniform}\left(0, \tfrac{1}{2}\right) \\
w_i &= z_x, \text{ where } x = \textstyle\sum_k \mathcal{I}\left(i \geq \tau_k\right) \\
\theta_i^{\mathrm{d}} &= \begin{cases} \mathcal{H}_{w_i}^{\mathrm{d}}\left(\boldsymbol{s}_i\right)\left(1-\epsilon^{\mathrm{d}}\right) + \tfrac{1}{2}\epsilon^{\mathrm{d}} & \text{if } w_i \in \{\text{ttb}, \text{tally}, \text{wadd}\} \\ \tfrac{1}{2} & \text{if } w_i = \text{guess} \end{cases} \\
\theta_{ik}^{\mathrm{s}} &= \begin{cases} \left(1-\epsilon^{\mathrm{s}}\right) & \text{if } w_i \in \{\text{ttb}, \text{tally}, \text{wadd}\}, \mathcal{H}_z^{\mathrm{s}}\left(\boldsymbol{s}_i\right) = 1 \\ \epsilon^{\mathrm{s}} & \text{if } w_i \in \{\text{ttb}, \text{tally}, \text{wadd}\}, \mathcal{H}_z^{\mathrm{s}}\left(\boldsymbol{s}_i\right) = 0 \\ \tfrac{1}{4} & \text{if } w_i = \text{guess} \end{cases} \\
\theta_{ik}^{\mathrm{r}} &= \begin{cases} \left(1-\epsilon^{\mathrm{r}}\right) & \text{if } \mathcal{H}_{w_i}^{\mathrm{r}}\left(\boldsymbol{s}_i\right) = 1 \\ \epsilon^{\mathrm{r}} & \text{if } \mathcal{H}_{w_i}^{\mathrm{r}}\left(\boldsymbol{s}_i\right) = 0 \\ \tfrac{1}{2} & \text{if } \mathcal{H}_{w_i}^{\mathrm{r}}\left(\boldsymbol{s}_i\right) = \tfrac{1}{2} \end{cases} \\
y_i^{\mathrm{d}} &\sim \text{Bernoulli}\left(\theta_i^{\mathrm{d}}\right) \\
y_{ik}^{\mathrm{s}} &\sim \text{Bernoulli}\left(\theta_{ik}^{\mathrm{s}}\right) \\
y_{ik}^{\mathrm{r}} &\sim \text{Bernoulli}\left(\theta_{ik}^{\mathrm{r}}\right)
\end{aligned}
$$

Figure 11. Graphical model for inferring strategy switches across trials, based on decision, search, and verbal report data.

panels show the inferred use of strategies, and switches between strategies, for three cases. The first case is when only decisions are considered, as for the graphical model in Figure 7 and already presented in Figure 8. The second case is when both decision and search data are considered. The graphical model required to make this inference is the same as that presented in Figure 11, with the components related to the verbal report data removed. The third case is when all three behavioral data sources are available and uses the full graphical model in Figure 11.

The results in Figure 12 make it clear that inferences about strategy use, including switching between strategies, depend on what behavioral data are incorporated. In particular, more fine-grained inferences are made when more sources of behavioral data are made available. Based only on the decisions made by this participant, the inference is that they used the weighted-additive strategy until a little over half-way through the experiment, and then switched to the simpler tally strategy. If the search data are incorporated together with the decision data, however, additional switches in strategy use are inferred. Specifically, two relatively short periods of use of the noncompensatory take-the-best strategy are now inferred. These blocks of trials correspond to ones in which the participant almost always searches until a discriminating cue is found, or one cue beyond that critical point. In other sequences of trials in the experiment, the search behavior of the participant typically involves many cues beyond the first discriminating one being considered. These sequences remain inferred as involving compensatory strategy use. The same strategy switching inferences are made when verbal report data are additionally incorporated.

Figure 13 summarizes the results of applying this model to Participant 1, using the same format. Once again, it is clear the inferences about strategy use depend on the available data. Based on only the decisions, the inference is that the participant uses the take-the-best strategy throughout the experiment. With the additional evidence provided by the search behavior, however, a brief switch is inferred, involving the use of a compensatory strategy—both the tally and weighted-additive strategies are equally possible—on Trials 52 and 53. On these two consecutive trials, the participant searched well beyond the first discriminating cue, which they rarely did on other trials. Coupled with the misprediction the take-the-best strategy makes about their decision on Trial 52, the search data lead to sufficient evidence to infer a switch in strategy.

Interestingly, the presence of this switch gains further support from the verbal report

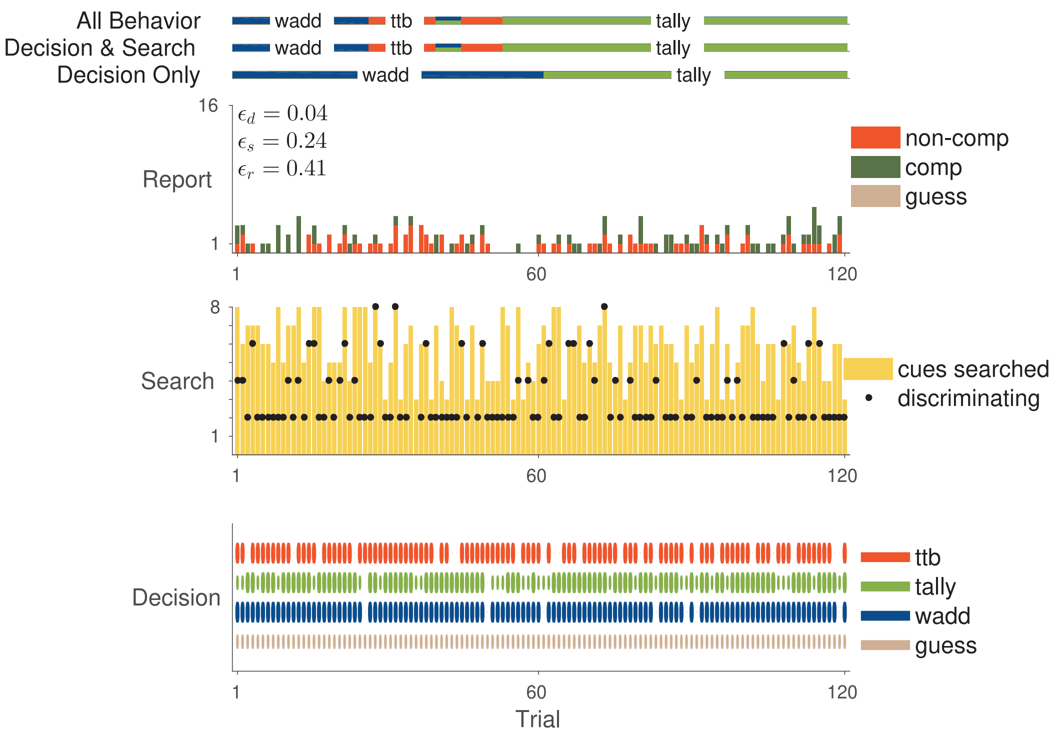


Figure 12. The decision, search, and verbal report behavior of Participant 8, and corresponding inferences about strategy use. The bottom panel shows the predictive accuracy for the take-the-best, tally, weighted-additive, and guessing strategies on each trial. The bars in the middle panel show the number of cues searched on each trial, and the filled circles indicate finding a discriminating cue. The bars in the top panel show the number of verbal reports consistent with compensatory and noncompensatory strategies for each trial. The horizontal bars above the verbal report panel show the inferred strategy, and changes in strategy, for the cases where the decision, decision and search, and all behaviors are used in inference. See the online article for the color version of this figure.

data. On Trials 51 and 52 the participant makes verbal reports consistent with a compensatory strategy.

> Trial 51: "Umm XPV is a yes for share trend positive. While VTW is a no, and XPV is also invest in new projects. So it's even higher percent, so I'm definitely going to choose that one."

On Trial 51, the two imaginary stocks are "XPV" and "VTW." The highest validity attribute, as randomly assigned at the beginning of the experiment, was "Share trend positive?" and the second highest validity attribute was "Invest in new projects?" Search data show the participant reveals the values of "Share trend positive?" for both stocks, and the verbal protocol data make clear the participant attends and encodes the attribute values (first two statements). XPV is a Yes, and VTW is a No. There is now, already, highest available validity discriminating evidence in favor of XPV. A strict interpretation of the take-the-best strategy predicts the participant stops there and selects XPV, but that is not what happens. Instead, the participant goes on to reveal attribute values for both stocks for "Invest in new projects?" and the verbal protocol includes two statements relevant to attending and encoding this additional information about the stocks. Thus, both the search data and the verbal protocol data provide evidence consistent with a compensatory strategy on this trial.

> Trial 52: "Umm JWB is a yes for share trend positive, while QDC is a no. Umm but QDC does invest in new projects. Umm and it also is established company. So I am going to choose QDC."

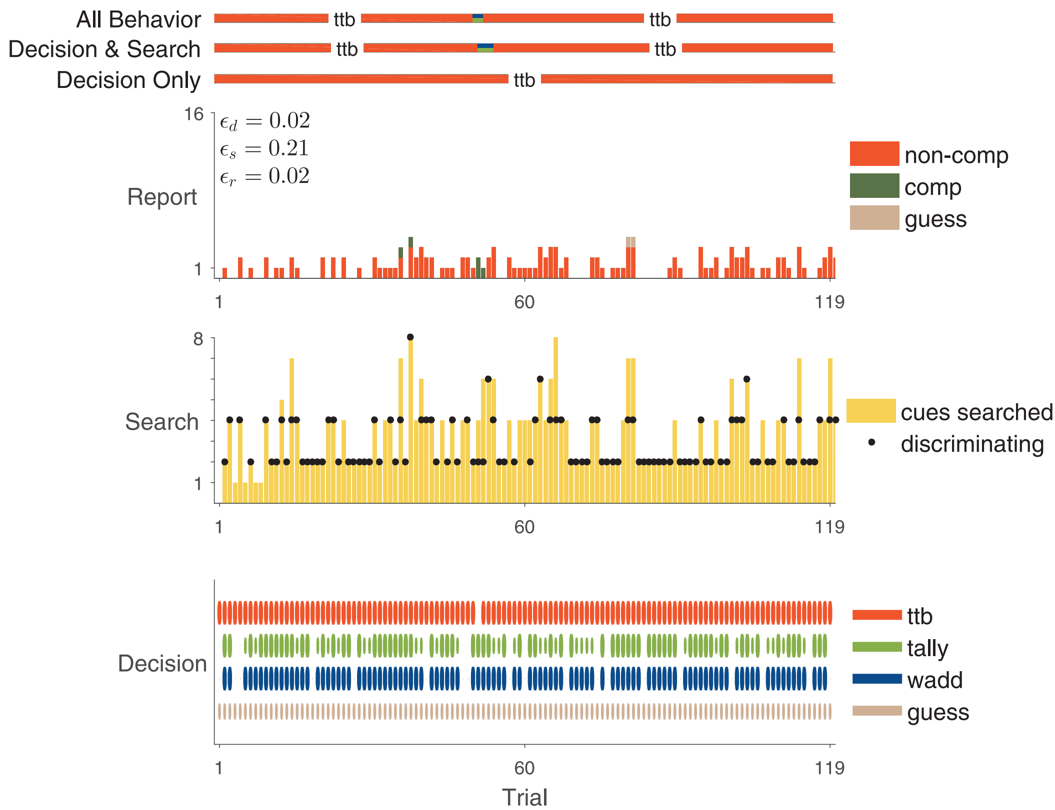


Figure 13. The decision, search, and verbal report behavior of Participant 1, and corresponding inferences about strategy use. The information presented takes exactly the same format as for Participant 8. See the online article for the color version of this figure.

On Trial 52, the two imaginary stocks are "JWB" and "QDC." It remains true that the highest validity attribute is "Share trend positive?" followed by "Invest in new projects?" and then "Established company?" Search data show the participant first reveals the value of "Share trend positive?" for both stocks, and the first two statements in the verbal protocol data make clear the participant attends and encodes the attribute values. JWB is a Yes, and QDC is a No. Once again, this is sufficient discriminating evidence if the participant is using the take-the-best strategy. However, the participant continues on, first revealing the attribute values for "Invest in new projects?" and verbalizing the encoding of that information in their third statement, and then revealing the attribute values for "Established company?" and verbalizing the encoding of that information in their fourth statement. The search data and the verbal protocol data both provide evidence consistent with a compensatory strategy on this trial.

As this analysis shows, on both Trials 51 and 52 both the search and verbalization data are consistent with a compensatory decision strategy. Accordingly, the inference of the model that incorporates the behavior is a brief switch on those trials from take-the-best to a compensatory strategy, before quickly returning to the take-the-best strategy.

Overall Results

Figure 14 shows the inferences made about switching strategy use, based on the decision, search, and verbal report data, for all participants. All but four of the participants are now inferred to change strategy at least once, and some, such as Participants 8 and 18, are inferred to switch strategy many times. The BFs in favor

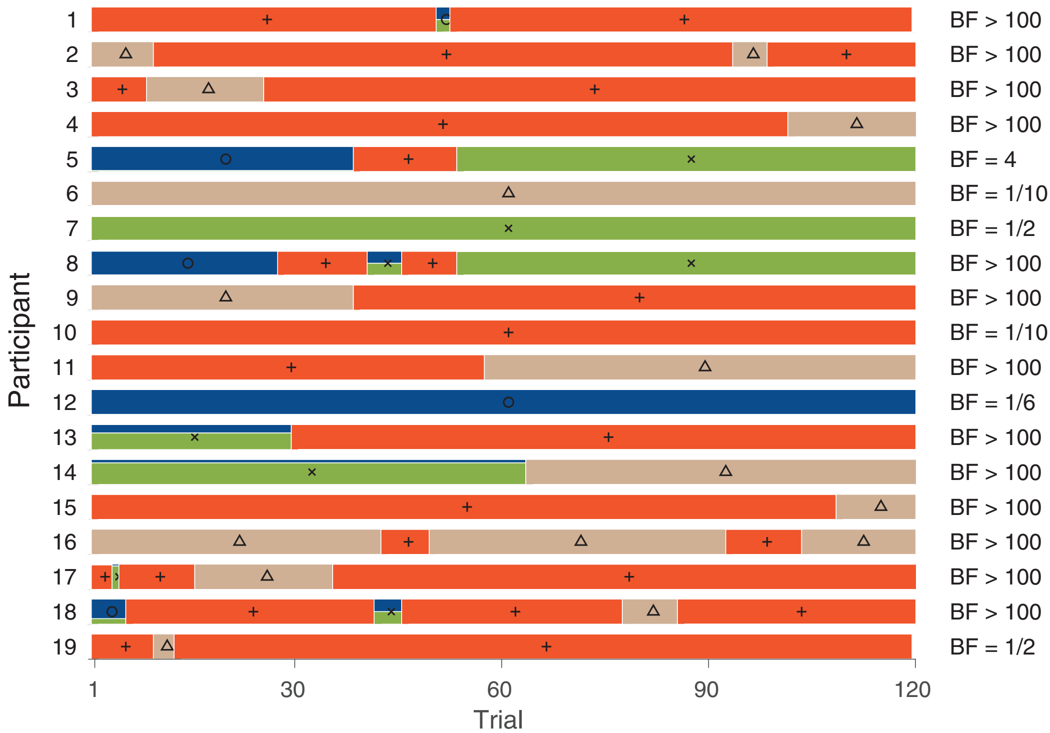


Figure 14. Inferences about switching strategy use for all 19 participants based on decision, search, and verbal report data. Red = take-the-best (+ marker); green = tally (× marker); blue = weighted additive (circle marker); beige = guess (triangle marker). See the online article for the color version of this figure.

of strategy switching compared to fixed strategy use are much more decisive than those reported in Figure 9 based only on decision data. For most of the participants for whom strategy switching is inferred, the behavioral data are more than 100 times more likely under the strategy switching model. The only exceptions are Participants 5 and 19, with less overwhelming BFs.[2]

Many of the inferences are best characterized as refinements, rather than differences, although Participant 6 remains an example of additional behavioral evidence leading to a different inference. The individual-level findings for Participant 8 and Participant 1 presented in the preceding text are typical of the sort of refinements seen for many other participants. The additional evidence provided by the three behavioral data sources allows brief periods of strategy switching to be inferred. The additional switches are usually compatible with all three behaviors, but all three are needed to provide enough evidence to justify the additional complexity of inferring a switch in strategy use.

Prediction Performance

Finally, we consider external evaluation for the 15 participants who were inferred to have at least one switch in strategy. These evaluations now both take the form of prediction tests. Specifically, they involve measuring the ability of the models to predict the next decision and the next search behavior, based on observing the behavioral data on the preceding trials. Generalization tests are obviously now not possible

[2] The results for Participant 19 provide an interesting example of the limitations of basing strategy switching inferences on the mode of the joint distribution of $\boldsymbol{\tau}$ because the BF provides weak evidence for no switch points, despite two being inferred from the mode.

because the sources of behavioral data are all observed.

Figure 15 shows the results of these two prediction tests, using the same approach previously presented in Figure 10. Once again, the top panels correspond to the prediction of decisions, and the bottom panels now correspond to the prediction of search behavior. In both cases, the strategy switch model almost always either matches or outperforms the fixed model. The superiority is clearer for the decisions than search, which may reflect a relatively greater misspecification of search behavior in our modeling, although both models perform well above chance for the participants on the search measure, which provides a form of absolute assessment.

Discussion

We set out to address two major limitations pervading previous research modeling strategy use in multiple-cue decision making. The first was the methodological restriction to decision data in making inferences about strategy use. One of the emphases of the multiple-cue decision making literature is that search is an important part of the decision-making process, and different strategies make different assumptions about the extent and nature of search. As cognitive scientists, our interests extend beyond which decisions were made and what information was accessed to make those decisions. We are also interested in when and how that information was used in the process of arriving at those decisions. Verbal protocols provide a source of data into those processes. Given that strategy use generates predictions about multiple sorts of behavior, it is natural and sensible to base inferences about strategy use on all of these behaviors.

The second limitation we tackled was the theoretical assumption that a participant uses the same strategy on all of their decision trials. While almost universally assumed, this seems to be a strong and unrealistic position. Many previous studies have examined what properties of a task—the structure of the cue environment, task requirements that emphasize speed or accuracy, costs associated with search, and so on—might influence the use of one strategy or another (e.g., Bröder & Schiffer, 2003b; Newell & Shanks, 2003; Newell et al., 2003). Asking these questions allows the possibility that strategy use is not an invariant property of a person, but changeable under at least some circumstances, and the empirical evidence is that there are often large individual differences in strategy use (e.g., Bergert & Nosofsky, 2007; Hilbig & Moshagen, 2014; Lee & Cummins, 2004; Newell & Lee, 2011). It is a small theoretical leap from that position to allow the possibility that a person might change the strategy they use over trials within a task, based on self-regulation, adaptation, or inherent exploration, even without any external trigger from the environment or task demands.

To address these two limitations, we developed a series of models, and applied them to data reported by Walsh and Gluck (2016). We started with the standard approach of inferring fixed strategy use based on decisions. We then added the capability to base inferences on multiple data sources, then the capability to infer strategy switches, and finally combined these capabilities. For the set of strategies we considered, we found strong evidence that people switch among them, especially when considering all of their decision, search, and verbal report data. We found that the additional empirical evidence gained by considering multiple data sources helps justify the extension of strategy use to incorporate strategy changes. More complicated models require more evidence to be inferentially justified. With the behavioral data available, the strategy change model infers that only a few participants use the same strategy over all of the decision trials, and many participants switch strategies multiple times.

Our results suggest a hidden complexity in decision making that is masked by not allowing for strategy switches and considering only decision data. The challenge for the future is to identify regularities in when and why people switch strategies and develop models of the learning and adaptation processes involved. One possible regularity is that shifts often replace a more complicated strategy with a simpler one. Of the three substantive strategies we considered, it is natural to consider the weighted-additive strategy as more complicated than the tally strategy which, in turn, is more complicated than the take-the-best strategy. There are several examples in the results presented in Figure 14 in which participants make one switch over the course of their decisions, from a

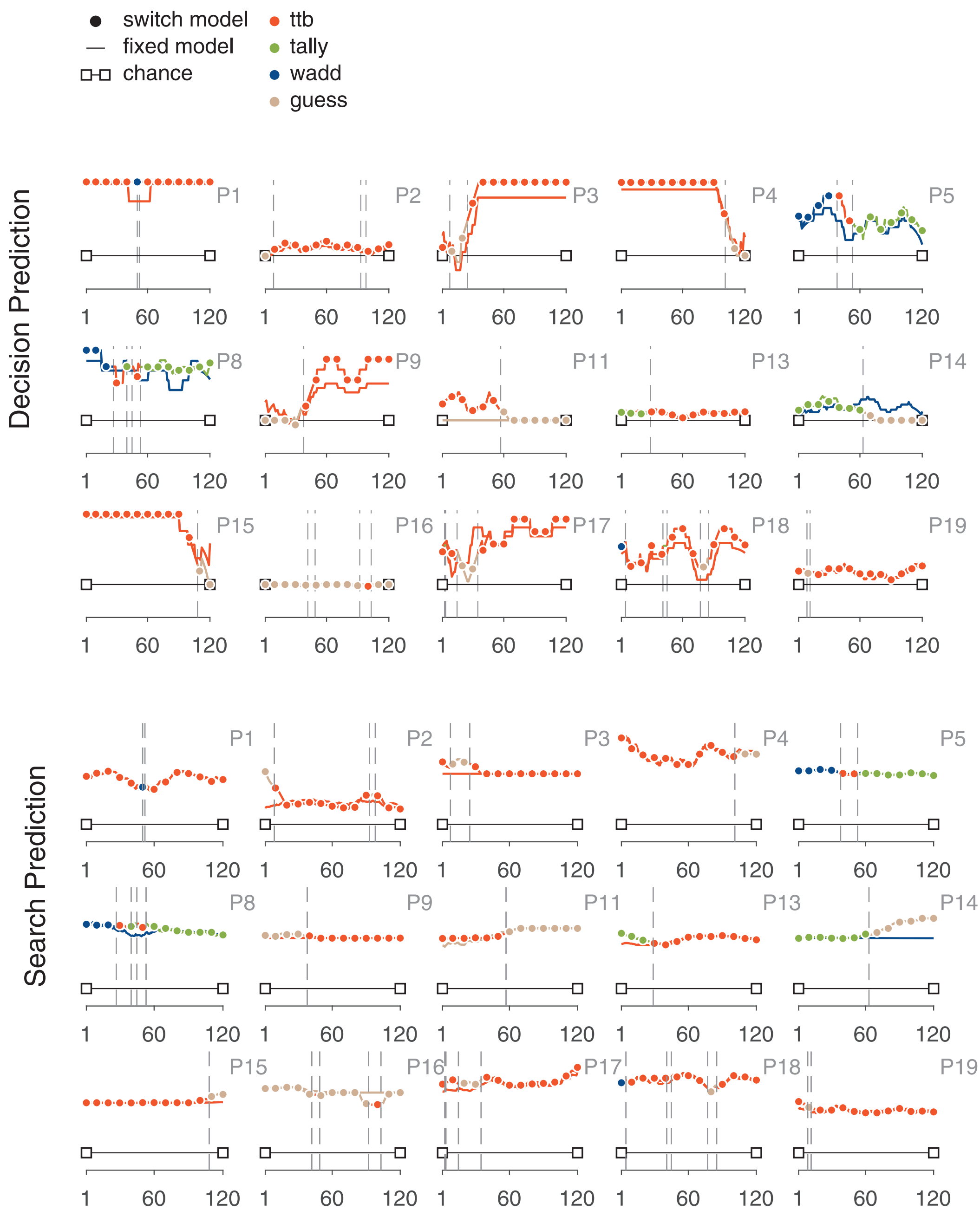


Figure 15. Predictive performance for decisions and search behavior for the model assuming switching strategies, the model assuming fixed strategies, and a chance model, based on decision, search, and report data. The top panels relate to the ability of the models to predict the decision on the next trial, based on observing all behavior up to the preceding trial. The bottom panels relate to the ability of the models to predict the search behavior on the next trial, based on observing all behavior up to the preceding trial. Individual panels correspond to participants, and lines show the smoothed predictive performance over the experimental trials. See the online article for the color version of this figure.

more complicated strategy to a less complicated one. There are no examples of a one-off switch to a more complicated strategy. A related observation is that guessing is common. This could be interpreted as some combination of inconsistent initial behavior or exploration while developing a decision strategy, indicative of search for a replacement strategy, and failing to execute a strategy to the completion of the task. It may be that these sorts of behavior are regularly observed when strategy switching is modeled, and processes like guessing are included among the candidate strategies. Beyond these initial suggestions, however, much more thorough investigation is needed to understand strategy switching. We considered only a single experiment that did not manipulate the decision environment or task demands and was not designed to relate individual differences in strategy use to relatively stable psychometric properties of individuals.

Besides examining additional empirical evidence, future work should expand the range of strategies considered. We focused on three central substantive strategies, the weighted-additive, tally, and take-the-best strategies, that are widely considered in the literature, but there are other possibilities. Hilbig and Moshagen (2014) have recently developed a probabilistic extension of the weighted-additive strategy, Heck et al. (2017) proposed an analogous probabilistic extension of the take-the-best strategy, and Oh et al. (2016) developed a "drop the worst" lexicographic strategy in which some of the least valid cues are not considered. Other natural extensions of the take-the-best strategy have also been considered, involving different mechanisms for search (Todd & Dieckmann, 2005), or requiring more than one discriminating cue to be found (Dhami, 2003). The sequential sampling framework has sometimes been proposed as a theoretical basis for unifying and systematizing some of these sorts of strategy variants (Lee & Cummins, 2004; Newell, 2005), and could be incorporated as a more heavily parameterized strategy (Lee & Zhang, 2012).

As these strategies are included in future analyses, it will be important to develop better models of search behavior. The deterministic take-the-best, tally, and weighted-additive strategies we considered provide almost a caricatured account of search. Strictly interpreted, they predict exactly one search pattern, which is almost never observed empirically. Our approach was to loosen the predictions by shifting the modeling focus to individual search behaviors, rather than the overall pattern, and allowing for execution error. We think this is a reasonable first approximation, but the prediction and generalization tests of search behavior suggest it is far from completely satisfactory. A better approach would be based on more theoretically motivated models of search that predict distributions of search behavior. Lieder and Griffiths (2017) have recently developed generative models of search, with rational underpinnings, that provide one promising approach. As part of the development of search models, the exact relationship between search and decision will need to be understood. There is interesting empirical evidence that this relationship can be subtle and complicated (Senter & Wedell, 1999). Perhaps higher density process data, such as verbal protocols and eye-tracking, or even neuro-functional data, will be informative in this regard.

We were also limited in our focus on sudden and unstructured switches between strategies as an account of change. It would be possible, for example, to incorporate theoretical assumptions about the transitions between strategies, such as making the transition from compensatory to noncompensatory strategies more likely. Our current model assumes no structure between transitions, with each different strategy being equally likely after each change point. Similarly, our model makes no theory-driven assumptions about how often or after how long people switch strategies. It could be especially productive to consider incorporating theories about both strategy transitions and durations so that, for example, short switches to and from a strategy became more likely for strategies like guessing.

Ultimately, the modeling goal is to move from inferring strategy change in available behavioral data to understanding and predicting change, using a model that incorporates theories of learning, adaptation, or self-regulation. The literature on multiple-cue decision making contains many ideas about possible learning mechanisms (e.g., Glöckner, Betsch, & Schindler, 2010; Todd & Dieckmann, 2005). There are, however, relatively few current models that provide psychological mechanisms for switches in strategy use. Rieskamp and Otto (2006) developed an influential and general account of strategy selection based on standard reinforcement learning mechanisms. Lee et al. (2014) considered similar reinforcement learning approaches, focusing on how the extent of search might change based on external signals like errors in decisions, or internal signals like the effort expended in search. They also considered a different theoretical class of mechanisms based on the self-regulation of confidence (Vickers, 1979), including hierarchical approaches that can produce sudden step changes in behavior as well as incremental adjustment.

Even simple decisions can be complicated, especially when considering how different people make them at different times. In this article, we have considered the possibility not just that different people use different strategies, but that the same person switches between different strategies over time. To allow for these possibilities, we developed generative models of strategy use that use Bayesian methods to infer the number and nature of the change points where switches in strategy occur. We also extended these models to incorporate multiple sorts of behavioral evidence, formalizing not just the predictions different strategies make about decision making, but also their predictions about search behavior and verbal reports. We found strong evidence for switches in strategy use, especially when inferences are based on the available behavioral evidence. This paints a more complicated but interpretable and more realistic picture of the way people use simple strategies to make sequences of decisions.

References

Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

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

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

Bobadilla-Suarez, S., & Love, B. C. (2018). Fast or frugal, but not both: Decision heuristics under time pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 24–33.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky choice with heuristics: Reply to Birnbaum (2008); Johnson, Schulte-Mecklenbeck, and Willemsen (2008); and Rieger and Wang (2008). *Psychological Review, 115,* 281–290.

Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as amodel of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1332–1346.

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

Bröder, A., & Schiffer, S. (2003b). "Take-the-best" versus simultaneous feature matching: Probabilistic inferencesfrom memory and the effects of representation format. *Journal of Experimental Psychology: General, 132,* 277–293.

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 904–918.

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171–189.

Cox, R. T. (1961). *The algebra of probable inference*. Baltimore, MD: Johns Hopkins University Press.

Dhami, M. K. (2003). Psychological model of professional decision-making. *Psychological Science, 14,* 175–180.

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

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing, 16,* 203–213.

Fechner, H. B., Schooler, L. J., & Pachur, T. (2018). Cognitive costs of decision-making strategies: A resource demand decomposition analysis with a cognitive architecture. *Cognition, 170,* 102–122.

Feynman, R. (1994). *The character of physical law*. New York, NY: Modern Press.

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

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

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

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

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

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making, 23,* 439–462.

Grünwald, P. D. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.

Heck, D. W., Hilbig, B. E., & Moshagen, M. (2017). From information processing to decisions: Formalizing and comparing psychologically plausible choice models. *Cognitive Psychology, 96,* 26–40.

Hilbig, B. E. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality, 42,* 1641–1645.

Hilbig, B. E., & Moshagen, M. (2014). Generalized outcome-based strategy classification: Comparing deterministic and probabilistic choice models. *Psychonomic Bulletin & Review, 21,* 1431–1443.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York, NY: Cambridge University Press.

Jekel, M., Nicklisch, A., & Glöckner, A. (2010). Implementation of the multiple-measure maximum likelihood strategy classification method in R: Addendum to Glöckner (2009) and practical guide for application. *Judgment and Decision Making, 5,* 54.

Jordan, M. I. (2004). Graphical models. *Statistical Science, 19,* 140–155.

Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica, 56,* 81–92.

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50,* 456–460.

Lee, M. D. (2004). An efficient method for the minimum description length evaluation of cognitive models. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 807–812). Mahwah, NJ: Lawrence Erlbaum.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55,* 1–7.

Lee, M. D. (2016). Bayesian outcome-based strategy classification. *Behavior Research Methods, 48,* 29–41.

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In J. Wixted & E.-J. Wagenmakers (Eds.), *The Stevens' handbook of experimental psychology and cognitive neuroscience, Vol. 5: Methodology* (4th ed., 37–84). New York, NY: Wiley.

Lee, M. D. (in press). A simple and flexible Bayesian method for inferring step changes in cognition. *Behavior Research Methods*.

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

Lee, M. D., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision Making, 6,* 832–842.

Lee, M. D., Newell, B. R., & Vandekerckhove, J. (2014). Modeling the adaptation of the termination of search in human decision making. *Decision, 1,* 223–251.

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin & Review, 25,* 114–127.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press.

Lee, M. D., & Zhang, S. (2012). Evaluating the process coherence of take-the-best in structured environments. *Judgment and Decision Making, 7,* 360–372.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review, 124,* 762.

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

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

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2015). *Straight choices: The psychology of decision making*. New York, NY: Psychology Press.

Newell, B. R., & Lee, M. D. (2011). The right tool for the job? Comparing evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making, 24,* 456–481.

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

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

Oh, H., Beck, J. M., Zhu, P., Sommer, M. A., Ferrari, S., & Egner, T. (2016). Satisficing in split-second decision making is characterized by strategic cue discounting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1937.

Pachur, T., Hertwig, R., Gigerenzer, G., & Brandstätter, E. (2013). Testing process predictions of models of risky choice: A quantitative model comparison approach. *Frontiers in Psychology, 4,* Article 646.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Pearl, J. (1998). Graphical models for probabilistic and causal reasoning. *Handbook of Defeasible Reasoning and Uncertainty Management Systems: Quantified Representation of Uncertainty and Imprecision, 1,* 367–389.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491.

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision, 2,* 237–279.

Rieskamp, J., & Otto, P. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135,* 207–236.

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

Roberts, S., & Pashler, H. (2002). Reply to Rodgers and Rowe (2002). *Psychological Review, 109,* 605.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review, 120,* 39.

Senter, S. M., & Wedell, D. H. (1999). Information presentation constraints and the adaptive decision maker hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 428.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika, 62,* 407–416.

Todd, P. M., & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 17, pp. 1393–1400). Cambridge, MA: MIT Press.

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science, 25,* 169–176.

Walsh, M. M., & Gluck, K. A. (2016). Verbalization of decision strategies in multiple-cue probabilistic inference. *Journal of Behavioral Decision Making, 29,* 78–91.

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E. (2010). An encompassing prior generalization of the Savage-Dickey density ratio test. *Computational Statistics and Data Analysis, 54,* 2094–2102.

Zeigenfuse, M. D., & Lee, M. D. (2010). A general latent assignment approach for modeling psychological contaminants. *Journal of Mathematical Psychology, 54,* 352–362.

(*Appendices follow*)

Appendix A

Inferring Strategy Switches

There is a large statistical literature on Bayesian methods for detecting change points (e.g., Adams & MacKay, 2007; Fearnhead, 2006; Smith, 1975). The method we use to infer the number and location of strategy switches is a natural extension of the approach developed and demonstrated by Lee (in press). The method involves summarizing the information in the joint posterior distribution over the $\boldsymbol{\tau}$ and z parameters, which expresses the relative probability over all possible numbers of switch points and their locations. We do this using the posterior mode of $\boldsymbol{\tau}^*$ as a point estimate. Those values of $\boldsymbol{\tau}^*$ that are not 1 correspond to the trials at which a switch is inferred to occur. The posterior distribution over the strategy used between each switch point is then the conditional posterior distribution of z on $\boldsymbol{\tau}^*$.

Figure A1 provides a concrete example of this inference process, for Participant 7, based only on their decision data. The bottom panel shows the predictive accuracy of the four strategies. The panel labeled $p(\tau_i')$ shows the marginal prior distribution on each of the $\gamma = 5$ possible switch-point parameters $\tau_1', \ldots, \tau_5'$, before the order constraint is imposed. As specified in the graphical model in Figure 7, half the prior probability is given to the first trial, and the remaining half of the probability is equally distributed over Trials 2 to 120.

The panels labeled $p(\tau_i \mid y)$ show the inferred marginal posterior distributions, for the order constrained parameters. Figure A1 shows that the posterior distributions for $\tau_1, \ldots, \tau_3$ have clear modes at 1, while τ_4 and τ_5 have modes at other trial numbers. The joint posterior mode is $\boldsymbol{\tau}^* = (1, 1, 1, 84, 95)$. This is (partially) visualized in the inset panel, which shows the joint marginal posterior distribution of (τ_4, τ_5). This is a useful summary of the full five-dimensional joint posterior, because almost all of the posterior probability—as the marginal distributions make clear—in the other three dimensions lie at Trial 1.

The inset panel shows uncertainty about the values of the switch points τ_4 and τ_5, corresponding to a range of possibilities more or less consistent with the data. In our analyses, we use the mode, and so two switch points are inferred, at trials 84 and 95. In principle, the posterior distribution for the strategies used before and after these switch-points is present in the posterior samples from the original inference, by conditioning on $\boldsymbol{\tau}^*$. In practice, to obtain a large enough number of relevant samples to estimate the posterior for z accurately, we re-run the inference on a version of the graphical model that assumes the number and location of the switch points. Every posterior sample for z is thus correctly conditioned and provides a computationally efficient approach to inferring the strategy-use inferences shown at the top of Figure A1.

To approximate the BFs between the fixed strategy and strategy switching models, we use the Savage-Dickey method (Wetzels, Grasman, & Wagenmakers, 2010). This method is applicable, because the fixed strategy model is a nested special case of the general strategy switching model corresponding to $\boldsymbol{\tau} = (1, \ldots, 1)$. The ratio between the posterior and prior mass at this point in the joint parameter space provides an approximation to the BF in favor of the strategy switching model.

(*Appendices continue*)

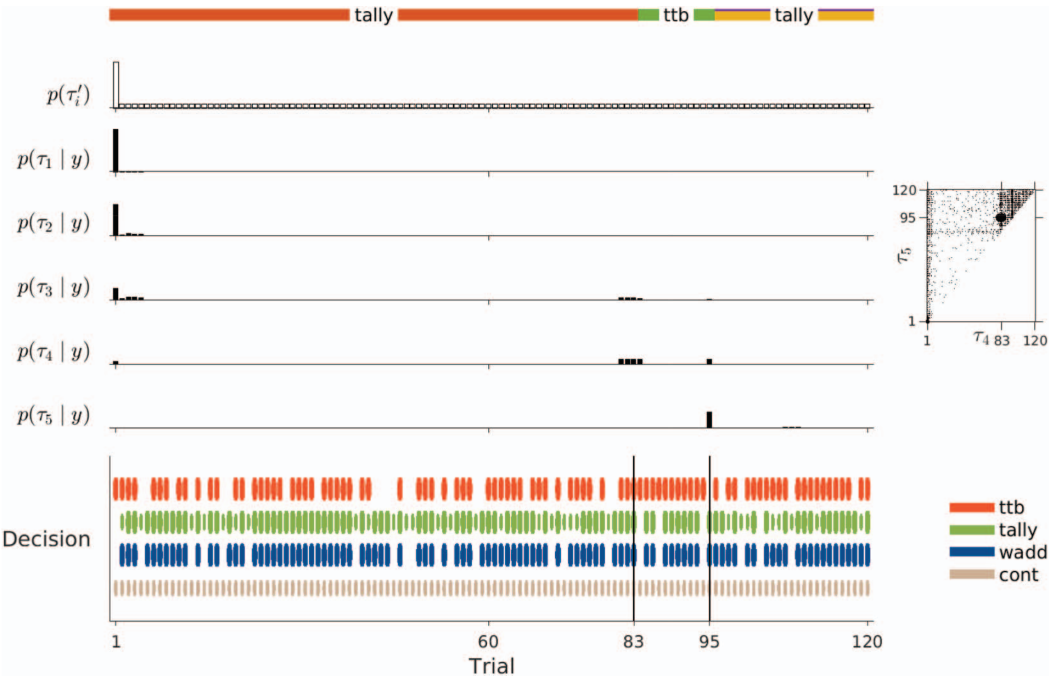


Figure A1. Demonstration of the approach to inferring strategy switches. The bottom panel shows the predictive accuracy for the take-the-best, tally, weighted-additive, and guessing strategies on each trial for Participant 7. The middle panels show the marginal posterior distribution for parameters $\tau_1, \ldots, \tau_5$ corresponding to five possible change points. The small inset panel shows the joint posterior distribution for τ_4 and τ_5 on which the inference of strategy switches at Trials 84 and 95 is based. The horizontal bar at the top shows the final inferences about switching strategy use. See the online article for the color version of this figure.

(*Appendices continue*)

Appendix B

Model Recovery Study

Model or parameter recovery studies are commonly reported in the cognitive modeling literature. In these studies, data are simulated by known models and parameter values, and evaluation hinges on the ability of a model or method to recover the ground truth. We think these studies are useful, but in a much more limited way than is sometimes asserted or implied (Lee, 2018). Simulation studies are useful as a way of providing a sanity check on the accuracy of model implementation, or for exploring the informativeness of experimental designs.

Model recovery studies are not useful, however, as a basis for making claims about the ability of the model to detect the "truth" from empirical data. The fact that a model can recover data simulated according to that model obviously does not provide evidence that the model corresponds to the data-generating processes that exist in nature. We think, in contrast, the sorts of prediction and generalization tests we presented provide good ways to assess the usefulness of a model as an account of how the world works.

Model recovery studies are also not useful for evaluating methods of inference themselves, in terms of testing how well a method infers the model or parameters used to generate data. This sort of evaluation is sometimes attempted (e.g., Pitt, Myung, & Zhang, 2002; Ratcliff & Childers, 2015), but we argue it confuses the concepts of *inference* and *inversion*. Inference finds what follows from the available information, which is the goal of model-based inferences in the empirical sciences. Inversion aims to recover the "truth" in a process more akin to deduction, as in the inversion of a matrix to solve a linear system of equations. Evaluating methods using model recovery capabilities amounts to evaluating methods for inference against benchmarks that only make sense in the context of inversion. The goal of inference methods in the empirical sciences is not to recover what "really" generated the data, but to say what does (and does not) follow from modeling assumptions and the available data. The models should make assumptions that try to capture reality, but inference on the models and their parameters should only aim to evaluate what follows from those assumptions. We see prediction and generalization tests, together with evaluation of substantive models relative to calibrating models like the guessing strategy, as better ways to assess the absolute adequacy of a model.

We believe using Bayesian methods is one approach, and perhaps the best approach, to ensure that model-based inference is correct. Bayesian methods provide a complete, coherent, and consistent approach to combining modeling assumptions with empirical data, following the rules of probability theory (Cox, 1961; Jaynes, 2003). If Bayesian methods are used correctly—that is, if the model and the analysis are coded correctly—then we argue that the inferences are automatically correct, given the model and data. To try to make these ideas concrete, we conducted a model recovery study based on simulated data.

Recovery Study

We conducted a recovery study involving 15 simulated participants. The task was yoked to the sequence of trials completed by the first 15 real participants from our experiment. For the first three simulated participants, it was assumed they used the same strategy on all trials. That strategy was randomly determined, with each strategy being equally likely. For the next three simulated participants, it was assumed they switched strategy once. The trial on which the switch occurred was chosen randomly, with each trial being equally likely, and the two

(*Appendices continue*)

strategies were also determined randomly, with each strategy being equally likely. The same approach was used for the remaining nine simulated participants, with three each having two, three, and four strategy switches, involving randomly-determined switch points and strategies before and after the switches. For each triple of simulated participants with the same number of switches, one was assumed to have an execution error $\epsilon_d^* = 0.05$, one was assumed to have $\epsilon_d^* = 0.10$, and one was assumed to have $\epsilon_d^* = 0.15$. This combination of the number of strategy switches, from 0 to 4, and the range of the execution error from 0.05 to 0.15, corresponds to the range of behavior inferred from the empirical data.

Overall Results

Figure B1 shows the results of applying our model, using the same inference methods applied to the human behavioral data. Each panel corresponds to one of the simulated subjects, with different numbers of "true" switch points arranged by rows (increasing from top to bottom), and different levels of execution error arranged by columns (increasing from left to right). Within a panel, the top bar shows the "true" strategy use, displaying in terms of the same colors used throughout the paper. The bottom bar shows the inferred strategy use. Figure B1 shows that, when just one strategy is used, without a switch in strategies, the inference matches the ground truth that generated the

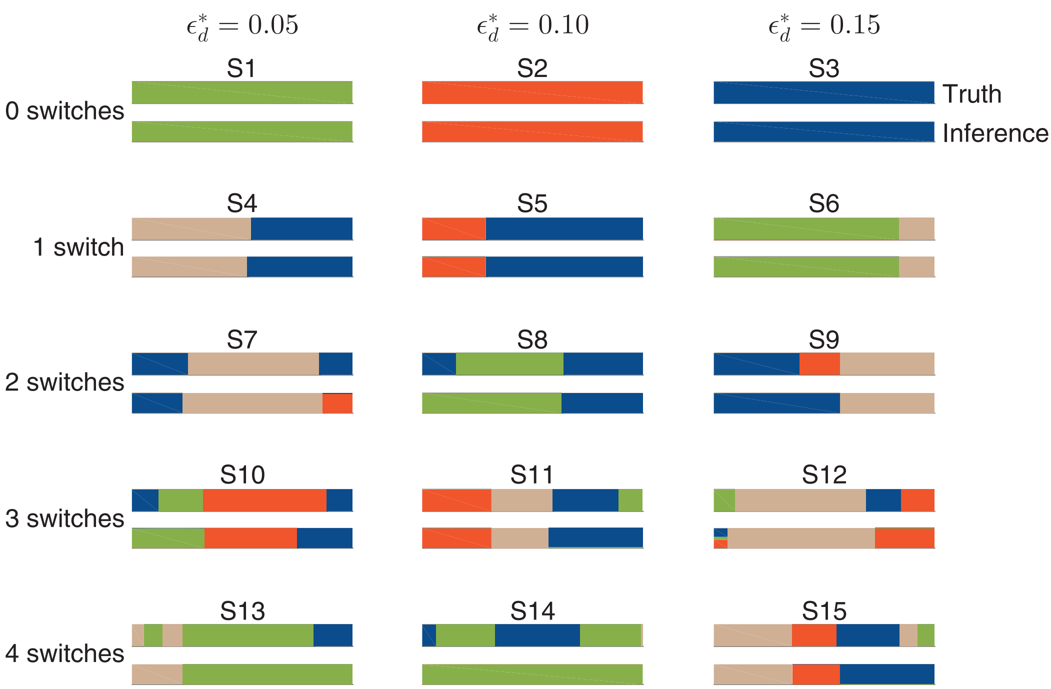


Figure B1. Results of a model recovery study, involving 15 different simulated strategy switching patterns. In each panel, the top bar corresponds to the ground truth, and the bottom bar corresponds to the inferred sequence of strategy use over trials. The panels are organized with different numbers of "true" strategy switches from 0 to 4 are in rows, and different levels of execution error $\epsilon_d^* = 0.05$, 0.10, and 0.15 are in columns. See the online article for the color version of this figure.

(*Appendices continue*)

data. As the number of switches used to generate the data increases, however, the inference is often simpler than the ground truth. For example, Simulated Participants S8 and S9 are inferred to have only one of the two "true" switches. This pattern of relatively simpler inferences is magnified for the simulated participants involving three or four switches.

Single-Subject Analyses

It is instructive to examine the details of the differences between inferences and the ground truths for individual simulated participants. Figure B2 shows the trial-by-trial predictions for each of the four strategies for Simulated Participant S4, together with the ground truth strategies that generated the simulated decision, and the modeling inferences. The inference approximately matches the ground truth, with a switch from the guessing to the weighted-additive strategy just over half-way through the experiment. But there is a difference in the exact trial on which this switch is inferred to take place. The ground truth is that the switch occurred at trial 66, whereas the inferred change is at Trial 63. Evaluating the inference by the standards of inversion—that is, testing whether the "true" and inferred trial agree—would result in finding a deficiency in the modeling.

We argue, however, that there is no deficiency in the inference, when it is evaluated as an inference. The details of Figure B2 make clear why. The simulated subject is "really" guessing on Trials 64, 65, and 66, but happens to make binary decisions that are completely consistent with the predictions of the weighted-additive strategy. Accordingly, the model infers that the strategy switch occurred three trials earlier than it "really" did. This inference is completely reasonable. Indeed, it would be strange and unreasonable if the inference matched the ground truth by which the data were simulated. The goal of inference is to say what follows from the data. As the concrete example in Figure B2 makes clear, the correct inference can be disconnected from the ground truth underlying simulated data.

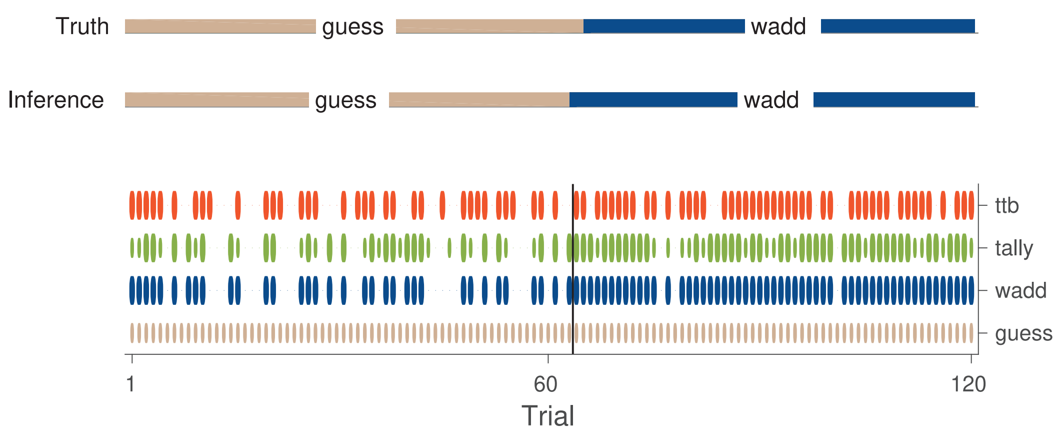


Figure B2. Results for the single Simulated Participant S4 from the model recovery study. The top bar shows the ground truth, which is that the participant guesses for 66 trials and then uses the weighted-additive strategy, with an execution error of $\epsilon_d^* = 0.05$. The bottom bar shows the model inference, which is that the participant guesses for 63 trials and then uses the weighted-additive strategy. Beneath the bars, the predictive accuracy of every possible strategy on each trial is shown, with the area of the markers corresponding to the agreement between strategy prediction and the simulated decision data. See the online article for the color version of this figure.

(*Appendices continue*)

While the difference between inference and inversion is small for Simulated Participant S4, the same disconnect plays out on a larger scale for other simulated subjects in the model recovery study and explains the differences between the inferences and ground truths apparent in Figure B1. For example, Simulated Participant S10 shows a larger difference in inferred and ground truth strategy use, involving disconnects not only in terms of when strategy switches occur, but how many switches there are, and what strategies are used. Figure B3 shows the details for this simulated subject. The ground truth is that there is an initial switch from the tally to the weighted-additive strategy, which the inference does not detect.

The details of the trial-by-trial predictions in Figure B3 make clear why. The tally strategy is equally as predictively accurate as the weighted-additive strategy for the initial trials that were "really" generated by the weighted-additive strategy. The simpler and correct inference is therefore that the tally strategy was used throughout. The inference about the switch from the tally to the take-the-best strategy roughly matches the ground truth, although the exact trial is slightly different, for exactly the sort of reasons discussed in relation to Simulated Participant S4 above. Finally, the inference is that the final transition from take-the-best to the weighted-additive strategy occurs much earlier than the ground truth. The cause is that, even though take-the-best is used to generate the decisions, errors in execution lead to a couple of trials for which simulated decisions are made that are inconsistent with take-the-best. These decisions happen to be predicted by the weighted-additive strategy, leading to the simple and sensible inference of weighted-additive strategy use for a long final set of trials.

The analysis of the other simulated participants summarized in Figure B1 leads to the same sort of findings. The inferences often do not agree with the ground truth, but are always reasonable, in the sense that they seem justified by the trial-by-trial patterns of agreement between the predictions of the strategies and the simulated decision data.

Conclusion

Our conclusion from the model recovery study is that it provides a sanity check on the accuracy of the implementation of the model. The inferences produced by the model make sense. The results also reinforce that the goal of

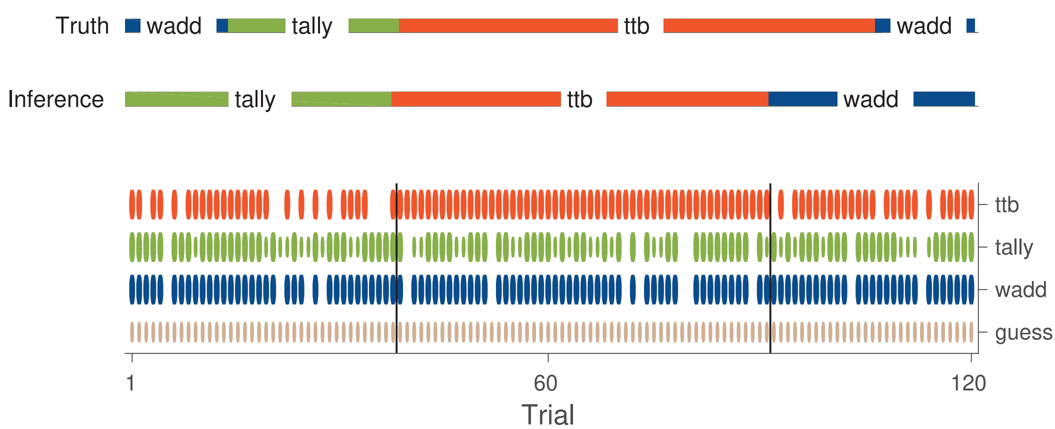


Figure B3. Results for the single Simulated Participant 10 from the model recovery study, using the same visual display as for Simulated Participant 4. See the online article for the color version of this figure.

(*Appendices continue*)

model-based inference is not to reveal the strategies that people "really" use, but to make inferences that are justified by the data, based on the assumptions made by the models. It is clear that inferences about strategy switching are often simpler than the ground truths used in the recovery study. That is reasonable, and a justification for considering multiple sources of behavioral evidence, since this additional evidence provides a basis for finer-grained inferences.

The discussion in the main body of this article of the different inferences obtained for real Participant 6—depending on whether only decision data are considered, or all behavioral data are considered—provides a good concrete example. As we argued, which of these inferences is "correct" is not a sensible statistical question. They are both correct, because they both follow from the assumptions made by the model, conditioned on the available data. Asking which inference is "correct" assumes that inversion is the right criterion for evaluation, which it is not. Asking whether an inference is correct really just amounts to a check on the accuracy of implementation of a model and the appropriateness of the use of Bayesian methods. To the extent that a model recovery study provides some reassurance on these issues, we think the results in Figure B1 are useful and positive.

Received January 24, 2018
Revision received December 18, 2018
Accepted January 4, 2019 ■