**The complexity of model-free and model-based learning strategies**

Alexandre L. S. Filipowicz[1,2,3], Jonathan Levine[1], Eugenio Piasini[3,4], Gaia Tavoni[3,4], Joseph W. Kable[2,3]*, Joshua I. Gold[1,3]*

Departments of Neuroscience[1], Psychology[2], and Physics[4], Computational Neuroscience Initiative[3], University of Pennsylvania
**co-senior authors*

**Corresponding author:**

Alexandre Filipowicz
D503 Richards Medical Research Building
3700 Hamilton Walk
University of Pennsylvania
Philadelphia, PA, 19104

Email: alsfilip@pennmedicine.upenn.edu

**Abstract**

A proposed continuum of learning strategies, from model-free to model-based, is thought to progress systematically in complexity and therefore flexibility. Here we distinguish different forms of complexity to show that, contrary to this idea, strategies at both ends of this continuum can be equally flexible. Using a canonical learning task, we first simulated behavior to show that computational complexity, a measure of implementation demands, is higher for a standard model-based versus model-free algorithm, but information complexity, a measure of flexibility, is not. We then analyzed human behavior to show that information complexity, which unlike computational complexity can be estimated from behavior, tended to increase for strategies that were increasingly either model-free or model-based, resulting in similar accuracy, suboptimal use of information, and increased response times. Thus, model-free and model-based strategies can have similar overall flexibility and instead are better distinguished by the specific task features from which they learn.

## Introduction

Humans can adopt a broad range of flexible strategies to learn from past observations to make accurate predictions about the future (Filipowicz, Anderson, & Danckert, 2016; Nassar, Wilson, Heasly, & Gold, 2010; O'Reilly, 2013; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). A prominent framework for capturing some of this range of learning strategies is based on a distinction between model-free and model-based reinforcement learning (Daw, Niv, & Dayan, 2005; Decker, Otto, Daw, & Hartley, 2016; Eppinger, Walter, Heekeren, & Li, 2013; Gläscher, Daw, Dayan, & O'Doherty, 2010; Kool, Cushman, & Gershman, 2016; Pauli, Cockburn, Pool, Pérez, & O'Doherty, 2018; Sutton & Barto, 1998). Model-free strategies learn action policies that aim to maximize reward using only past stimulus-action outcomes. These strategies have been equated with low-level habit formation and are thought to be relatively automatic. In contrast, model-based strategies learn action policies using both past stimulus-action outcomes and explicit representations of learned, latent statistical properties of the task (e.g., transition probabilities between different states). These strategies have been equated with more deliberate and flexible goal-directed learning (Daw et al., 2005; Gillan, Otto, Phelps, & Daw, 2015; Gläscher et al., 2010; Kool, Gershman, & Cushman, 2018; Kool, Gershman, & Cushman, 2017; Pauli et al., 2018). In general, people tend to use a mix of model-free and model-based strategies, with individual differences that can depend on the specific task conditions (Kim, Park, O'Doherty, & Lee, 2018; Kool et al., 2017), normal development (Decker et al., 2016; Eppinger et al., 2013), and certain psychiatric conditions (Sebold et al., 2014; Voon et al., 2015).

In computer science, a key distinguishing feature between model-free and model-based algorithms is their computational complexity. Specifically, model-free algorithms are typically less computationally complex than model-based algorithms, in terms of the number of operations required to implement the algorithm and how this number scales with more input. This language has begun to permeate the human learning and decision-making literature, such that model-based strategies are often described as having more strategic 'sophistication' or 'complexity' than model-free strategies (Decker et al., 2016; Doll, Simon, & Daw, 2012; Kim, Park, O'Doherty, & Lee, 2018; Kool et al., 2018). In some cases, these claims have focused on an analogous form of computational complexity as from computer science, which in the brain might reflect how many cognitive, metabolic, and/or other resources are used to perform a particular task (Bossaerts, Yadav, & Murawski, 2019; Polonio, Di Guida, & Coricelli, 2015; Tavoni, Doi, Pizzica, Balasubramanian, & Gold, 2019). However, assessing the computational complexity of human

learning can be challenging, because it requires assumptions about the underlying algorithm and how it is implemented by the brain. Moreover, the computational complexity of a given learning strategy does not necessarily determine the flexibility with which that strategy processes new inputs and statistical contexts.

Strategic flexibility, which is more closely related to the automatic versus deliberative dichotomy often ascribed to model-free versus model-based learning, can be measured using a different form of complexity known as statistical complexity. Statistical complexity quantifies the size of the hypothesis or feature space over which a model performs its inference, such that models exploring larger spaces are considered more complex (Bialek, Nemenman, & Tishby, 2001; Grünwald & Rissanen, 2007; Myung, Balasubramanian, & Pitt, 2000). In general, models with more statistical complexity make use of a wider range of past observations to adjust future predictions but also risk oversensitivity to noise that can reduce predictive accuracy (i.e., the bias-variance tradeoff; Hastie, Tibshirani, & Friedman, 2001). Like computational complexity, statistical complexity is defined formally with respect to a specific algorithm (Myung et al., 2000), which for human subjects is typically not known exactly. However, unlike computational complexity, quantities related to statistical complexity can also be estimated non-parametrically from behavior using information-theoretic measures. These measures, which can be estimated using the information-bottleneck method (Tishby, Pereira, & Bialek, 2000), define information complexity as the information transfer between past observations and future behavior (Filipowicz, Glaze, Kable, & Gold, 2019; Gilad-Bachrach, Navot, & Tishby, 2003; Glaze, Filipowicz, Kable, Balasubramanian, & Gold, 2018). Information complexity increases for strategies that make more flexible use of past observations to guide behavior.

The goal of this study was to better understand how model-free and model-based strategies relate to both computational and information complexity in the context of a canonical adaptive learning task (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). We first used simulations to show that model-free strategies can have lower computational complexity but higher information complexity than model-based strategies. We then used the information-bottleneck method to estimate information complexity from behavioral choice data of human subjects. We show that information complexity tended to be greatest for subjects that used strategies that were primarily either model-free or model-based. The more-complex strategies had similar features, including higher accuracy, less optimal/efficient use of past information, and

longer response times (RTs) than less-complex strategies, but differed in the exact nature of the information being processed for model-free versus model-based strategies. Our results imply that the degree to which a learning strategy is more model-free or model-based is not a reliable indication of its information complexity or strategic flexibility but rather characterizes the feature space over which inference is performed: observable features for model-free learning and latent features for model-based learning.

**Results**

We assessed the complexity of mixtures of model-based and model-free strategies used by both simulated and human subjects performing a commonly studied two-step task (Daw et al., 2011). For this task, subjects make one of two first-step choices that each lead stochastically to one of two second-step states. Each first-step choice leads to one of the two second-step states with 0.7 probability (common transition) and the other state with 0.3 probability (rare transition; Fig. 1a). Subjects then choose between two additional actions in this second state, which lead stochastically to a reward. Reward probabilities for each second-step action drift independently according to a Gaussian random walk ($\mu$=0, $\sigma$=0.025) with reflecting boundaries at 0.25 and 0.75 (Fig. 1b). A subject's goal in this task is to accumulate as much reward as possible.

This task was designed to identify propensities towards more model-based or model-free strategies (see Kool et al., 2016 for review). Purely model-free strategies rely solely on the observed stimulus-action mappings without attempting to learn or use latent information about the transition between the first and second steps. As a result, model-free strategies tend to repeat first-step actions that lead to second-step reward, regardless of whether the first to second-step transition was rare or common. This tendency can be measured behaviorally as a main effect of reward on the probability of repeating a previous first-step action (Fig. 1c). In contrast, model-based strategies account for the latent transition structure of the task environment and therefore select first-step actions that maximize the chance of returning to a rewarding second-step state. This tendency can be measured behaviorally by the interaction of reward and transition type (rare or common) on the probability of repeating a first-step action (Fig. 1c). These two tendencies are modulated parametrically by the degree to which a standard learning algorithm is more model-based or more model-free, with increases in model-based behavior corresponding to systematically decreasing main effects of reward and systematically increasing interactions between reward and transition type (Fig. 1d).

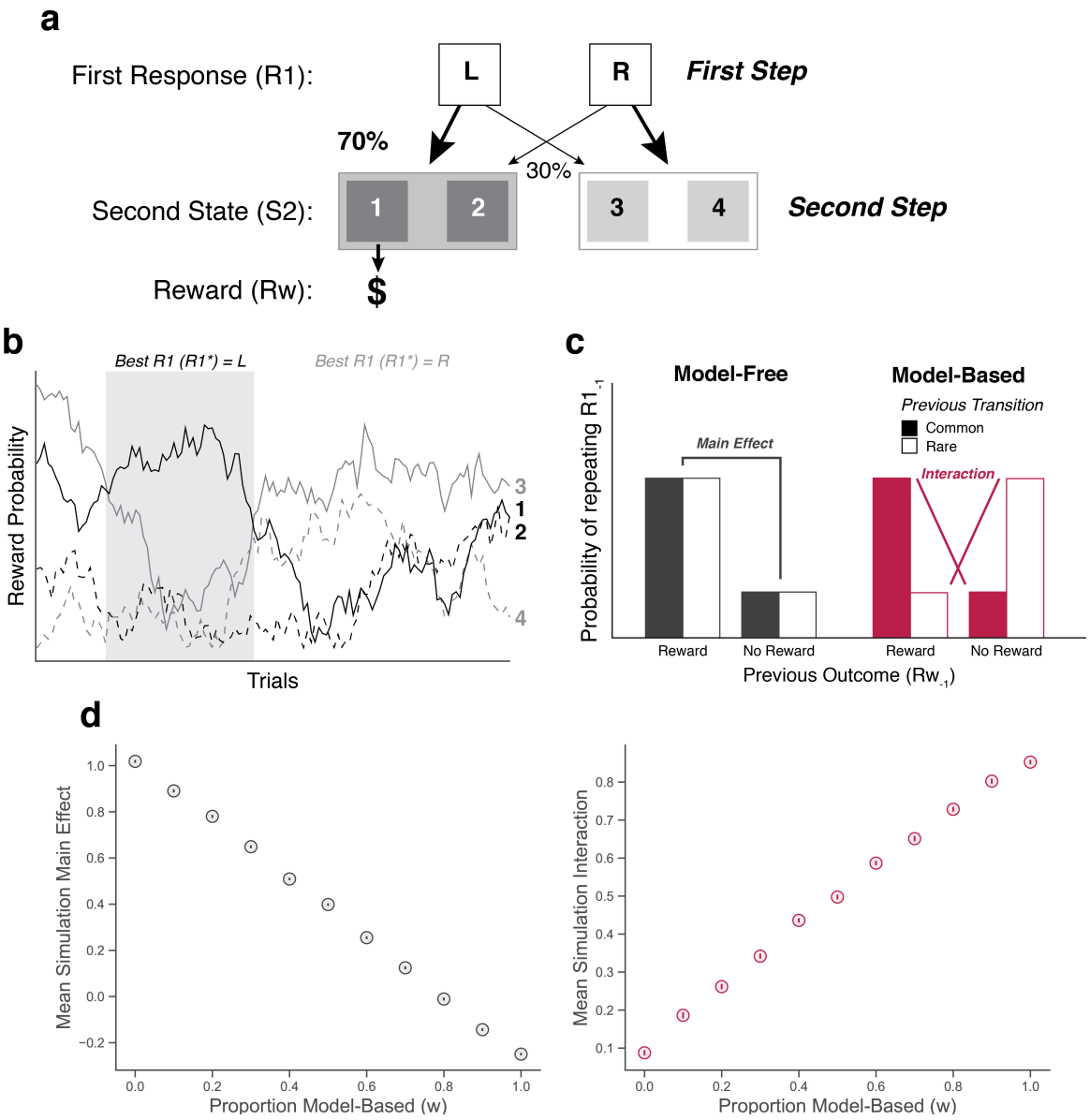


*Figure 1.* Description of the two-step task. (a) The subject makes a first-step choice between two alternatives (L or R) that lead probabilistically to one of two second-step states (filled or open rectangles). Subjects then make a second choice that leads probabilistically to a reward. (b) Example reward probabilities for each second-step option, which drift independently according to a Gaussian random walk. (c-d) Examples of response patterns for model-free and model-based strategies. (c) Model-free strategies tend to repeat first-step choices that lead to a reward, ignoring the transition structure of the environment. A behavioral signature of these strategies is a main effect of reward on the probability of repeating the same first-step choice as the previous trial. Model-based strategies use the transition structure of the environment to make choices. A behavioral signature of model-based strategies is an interaction between reward and previous transition type (common or rare) on the probability of repeating the same first-step choice as the previous trial. (d) Increased reliance on model-based strategies decreases the main effect of reward and increases the previous reward by transition type interaction.

*Computational and information complexity relate differently to different learning strategies*

To show that computational and information complexity can be dissociated for this task, we computed both quantities for a commonly used learning model that can simulate choice data based on a mixture of model-free and model-based strategies (Kool et al., 2016). This model computes both model-free and model-based solutions and makes choices by mixing those solutions with a mixture parameter $w$, such that $w = 0$ indicates a fully model-free strategy, $0 < w < 1$ indicates a mixed strategy, and $w = 1$ indicates a fully model-based strategy.

Because in this case we know the model-free and model-based learning algorithms, we can compute and compare their computational complexity directly. We defined computational complexity as the number of operations each algorithm uses to produce its responses, such that addition, subtraction, multiplication, and argmax each count as one operation. Other approaches that take into account different computational costs of each operation (e.g., multiplication and argmax might require more computations for the brain to perform than operations such as addition or subtraction; Eliasmith & Anderson, 2003) can yield results that differ quantitatively but show the same qualitative trends. These model-free and model-based algorithms differ only in the computations they use to derive action values for first-step responses. We therefore compared computational complexity in terms of the number of operations used to compute first-step action values by the model-free versus model-based algorithm.

Using this approach, pure model-based strategies, which required 10 operations to derive first-stage action values, were more computationally complex than pure model-free strategies, which required 6 operations (Fig. 2a,b). All mixture models with $0 < w < 1$ first compute model-free and model-based solutions separately and then mix those solutions according to the value of $w$. Under these conditions, all mixed strategies have the same computational complexity, equal to the sum of the number of operations needed to compute both the model-free and model-based solution plus additional operations to compute the mixture. Alternatively, a mixture strategy that uses either a model-free or model-based update rule on each trial, selected according to the mixture parameter $w$, would have, on average, a systematic increase in computational complexity from purely model-free to purely model-based.

We next measured the information complexity of different mixtures of model-free and model-based strategies using an information-bottleneck approach (Fig. 2c; Filipowicz et al., 2019;

Gilad-Bachrach, Navot, & Tishby, 2003; Glaze et al., 2018; Tishby, Pereira, & Bialek, 2000). This approach assumes that the subject' s strategy, or mental model ($M$), compresses information from the past ($X_{past}$) to predict future observations ($X_{future}$). Because the mental model cannot be observed directly, the behavioral responses ($R$) are used as a proxy for $M$ (Filipowicz et al., 2019; Glaze et al., 2018). We measured the size of the compression of past information ($I_{past}$) as the mutual information between past observations $X_{past}$ and $R$: $I_{past} = I(X_{past}; R)$. Although this quantity does not provide any details about the specific strategy or mental model being used to perform inference, it captures the amount of past information being encoded by the strategy and serves as a measure of its information complexity (Gilad-Bachrach et al., 2003). We also measured how effectively a particular strategy uses the information it encodes from the past as the mutual information between $R$ and future task elements that the model is attempting to predict, $X_{future}$: $I_{future} = I(R; X_{future})$ (Gilad-Bachrach et al., 2003; Tishby et al., 2000).

In the context of the two-step task, we defined past observations ($X_{past}$) as a feature vector ($f$) consisting of several task elements that the models could be using to perform their inference: 1) the first response ($R1$); 2) the second-step state into which subjects were transitioned ($S2$); 3) second-step feedback (reward or no reward; $Rw$); and 4) the latent best first response, which was defined as the response that would most likely lead to the second-step state with the most-rewarding option ($R^*$; Figs. 1b, 2c). Given the Markovian nature of the task environment and the strategies we implemented, we used only these features from the previous trial (Filipowicz et al., 2019; Glaze et al., 2018; Palmer, Marre, Berry, & Bialek, 2015). For each simulation, we measured information complexity as the mutual information between simulated responses and the feature vector observed on the previous trial. We measured predictive accuracy as the mutual information between simulated responses on the current trial and the best responses, given the actual reward probabilities on that trial ($R1^*$). We compared the relationship between information complexity and predictive accuracy to an empirical approximation of the optimal information bottleneck for this task, which describes the maximum predictive accuracy possible for a given degree of information complexity (Fig. 2d).

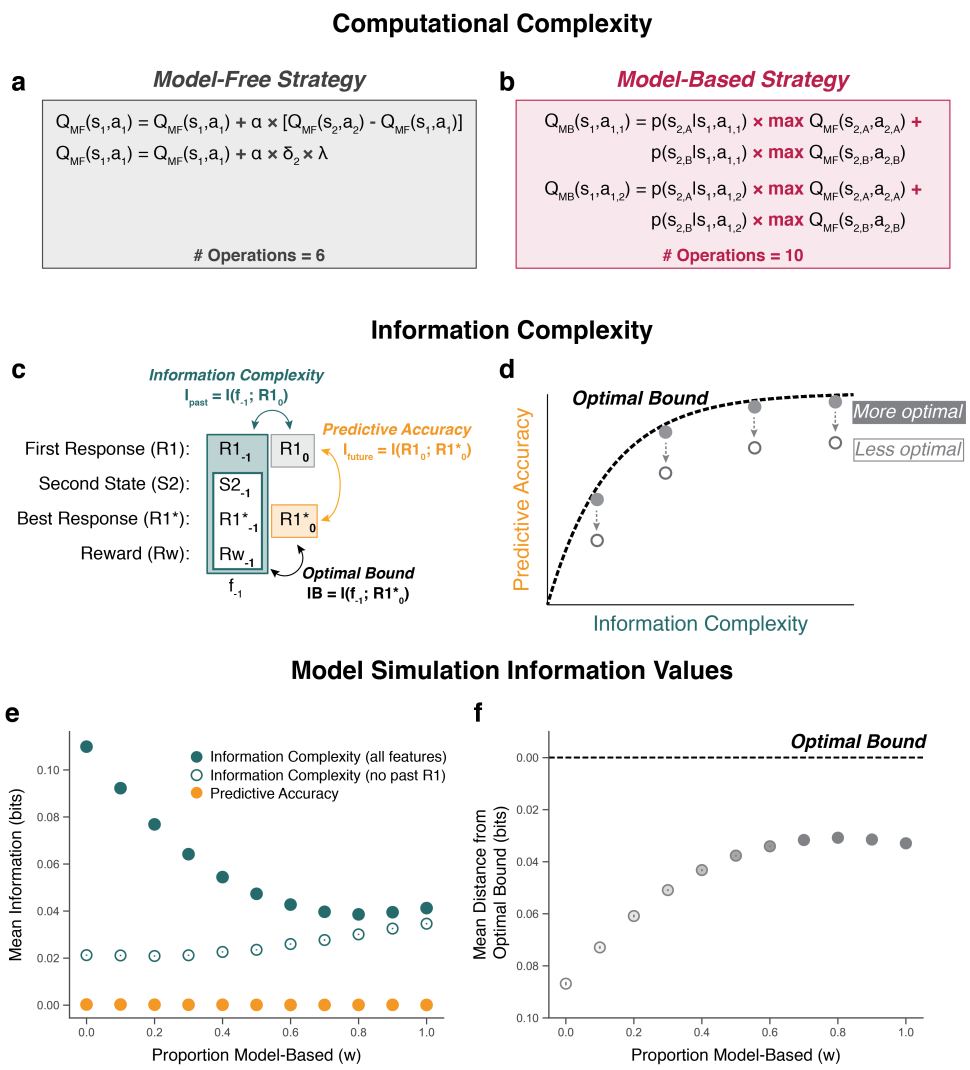


*Figure 2.* Computational and information complexity of model-free and model-based strategies. Computational complexity was measured as the number of operations implemented by a fully model-free (a) or fully model-based (b) strategy. Because these models differ only in how they compute first-step Q-values, only those equations were used to compare computational complexity (see methods for full equations). (c) Information complexity was measured as the mutual information between task features one trial in the past ($f_{-1}$, including first-step responses, $R1_{-1}$, second-step transitions, $S2_{-1}$, reward received, $Rw_{-1}$, and best first responses, $R1^*_{-1}$) and current responses ($R1_0$). Predictive accuracy was measured as the mutual information between $R1_0$ and $R1^*_0$. The optimal bound was calculated using an empirical approximation of the information bottleneck between the past feature vector and $R1^*_0$. (d) Visualization of the empirical information bottleneck, which shows the best possible predictive accuracy for a given information complexity. Points that fall below the bound are less optimal. (e) Mean information complexity (filled green circles), information complexity omitting $R1_{-1}$ from the past feature vector (open green circles), and predictive accuracy (orange circles) as a function of proportion model-based (*w* parameter) across 100 simulations of 10,000 trials. (f) Mean distance from the optimal bound as a function of proportion model-based strategies. Small lines inside open circles represent ±SEM, which is smaller than the size of the points in (e) and (f).

Using this metric, the average information complexity across 100 simulations was higher for pure model-free versus pure model-based strategies (mean information complexity: model-free=0.11 bits, model-based=0.04 bits; Wilcoxon rank-sum test comparing information complexity between purely model-free and purely model-based simulations, $p$=2.5x10$^{-34}$). For mixture strategies, there was a smooth, but non-monotonic, progression between those two endpoints, with a minimum at $w$=0.8 (Fig. 2e, closed circles). Across this range of mixtures, predictive accuracy was slightly higher for model-free over model-based strategies (mean predictive accuracy: pure model-free=0.0003 bits, pure model-based=0.0001 bits; Wilcoxon rank-sum test, $p$=0.0004), but this increase in predictive accuracy was much smaller that the relative increase in complexity for the model-free strategy. As a result, simulations of primarily model-based strategies were closer to the optimal bound than primarily model-free strategies, using fewer bits of information from the past to achieve similar levels of predictive accuracy (Fig. 2f).

These differences in information complexity could be attributed primarily to the flexibility provided by the reliance of model-free, but not model-based, strategies on the previous first-step response. The pure model-free strategy makes first-step choices according to action values that depend on whether the first action from the previous trial resulted in reward, regardless of whether this action led to a common or rare second-step transition. Conversely, the pure model-based strategy ignores its first-step response, considering only the second-step state into which it transitioned. This strategy then chooses the first-step action that most likely leads to the same second-step state if that second-step state was rewarded in the previous trial, or the opposite first-step action if that second-step state was not previously rewarded. Consistent with this interpretation, computing the mutual information between model-free responses and a past feature vector that omitted the previous first-step response resulted in a substantial decrease in information complexity (Fig. 2c, open circles).

Together, these simulations illustrate three key features of how canonical learning models relate to two different forms of complexity. First, pure model-free strategies can be less computationally complex than pure model-based strategies. Second, primarily model-free strategies can have higher information complexity than primarily model-based strategies. Third, such differences in information complexity depend critically on the specific task features each strategy uses to learn from the past and guide future choices, which may or may not be task relevant and therefore can affect the optimality of the learning process.

*Subjects using more model-free or model-based strategies have higher information complexity*

Because we do not have access to the exact algorithms used by human subjects to perform the task, we cannot assess their computational complexity precisely. However, we can assess their information complexity using the information-bottleneck approach. We used this approach to measure information complexity and predictive accuracy as a function of the degree of model-free versus model-based strategies from behavioral data of 197 subjects who performed the two-step task in a previous study (Kool et al., 2016). We used two measures to assess the degree to which subjects used model-based or model-free strategies. First, we fit behavioral data using the mixture model used for our simulations, leaving the mixture term $w$ as a free parameter. Second, we measured for each subject the main effect of reward on the probability of repeating the same first step action as the previous trial, as a metric of model-free tendencies, and the interaction between reward and transition type (rare or common), as a metric of model-based tendencies (Fig. 1c).

Contrary to the predictions from our simulations, there was no correlation between the mixture parameter ($w$) and either information complexity or predictive accuracy across subjects (Fig. 3a). There also was no correlation between the mixture parameter and how optimally subjects used the information they were encoding from the task (Fig. 3b) or their mean response times (Fig. 3c).

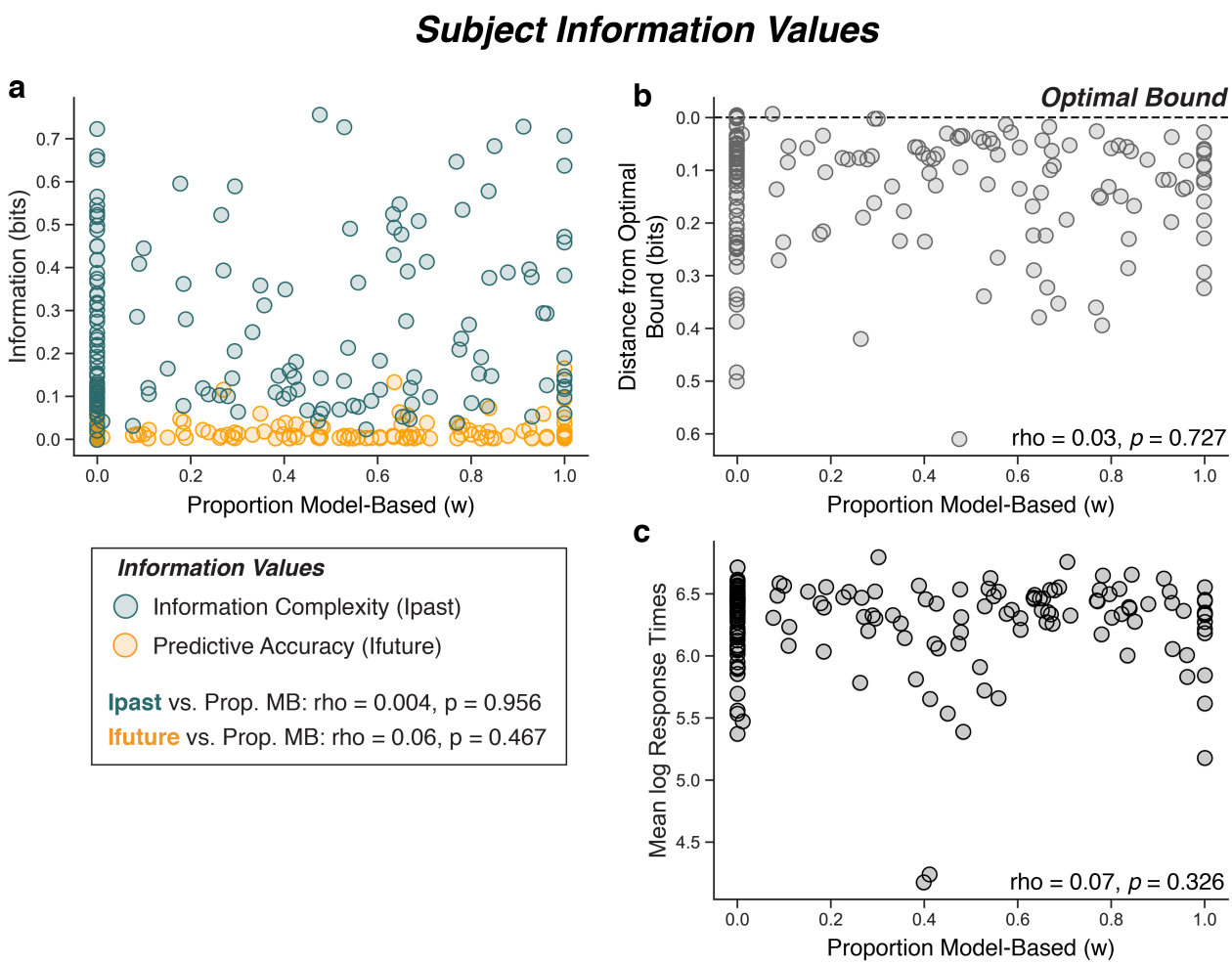


*Figure 3*. Comparison of the per-subject (circles) proportion model-based (MB) responses (best-fitting *w* parameter in the mixture model) and: (a) information complexity (green) or predictive accuracy (orange), (b) distance from the optimal information-bottleneck bound, and (c) response times. *rho* indicates Spearman's correlation coefficient; *p*-values are for $H_0$: *rho*=0.

However, there were systematic relationships between information complexity and the behavioral metrics that separately describe model-free versus model-based tendencies (the main effect and interaction terms, respectively, in Fig. 1c). Because the main effect and interaction terms were negatively correlated with each other (Spearman's rho = -0.24, $p$ = 0.002), we computed semi-partial correlations between complexity and each variable (main effect and interaction), while accounting for the other variable. As detailed below, these semi-partial correlations indicated increases in complexity for both model-free and model-based tendencies. Because the mixture parameter ($w$) is correlated positively with model-free tendencies (main effect) and correlated negatively with model-based tendencies (interaction), this result also accounts for the lack of systematic effects seen in Fig. 3.

Subjects with higher information complexity tended to have higher main effects (semi-partial Spearman's *rho*=0.32, $p$=0.00002; Fig. 4a), consistent with our simulations that showed an increase of information complexity for more model-free strategies (Fig. 2e). Like in our simulations, this correlation between main effect and information complexity was also reduced by removing the first response on the previous trial from the past feature vector (*rho* = 0.18; paired bootstrapped $t$-test for $H_0$: no decrease in correlation, $p$=0.007). The correlation between main effect and information complexity was also reduced when the reward from the previous trial was removed from the past feature vector (*rho*=0.26, $p$=0.006). This result confirms that subjects used trial-by-trial rewards to make future inferences. Removing any of the other elements from the past feature vector (second-step transition or best first-step responses) did not affect the correlation between main effect and information complexity (all $p$s>0.838).

Subjects with higher information complexity also tended to have higher interactions (*rho*=0.39, $p$=1.5x10$^{-7}$; Fig. 4b). This result is not consistent with our simulations that showed a decrease of information complexity for more model-based strategies (Fig. 2e). This correlation was decreased when reward was removed from the past feature vector (*rho*=0.32, $p$=0.02) but not affected by removing any other past elements from the past feature vector (first-step response, second-step transition, or best first-step response; all $p$s > 0.129).

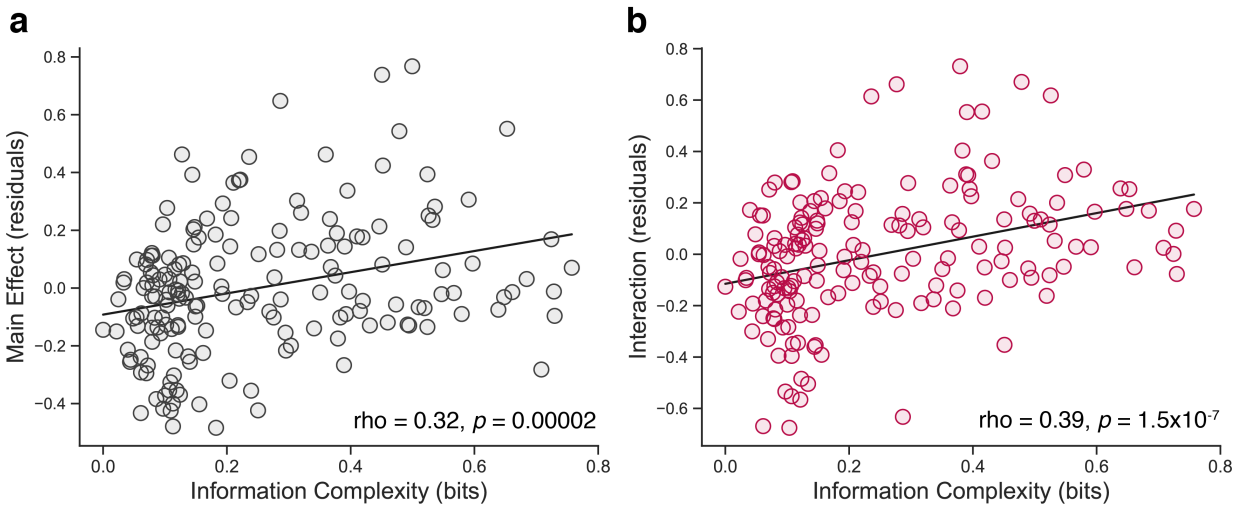


*Figure 4.* Information complexity was correlated positively with the size of both (a) main effect (a measure of the tendency to use a model-free strategy) and (b) interaction (a measure of the tendency to use a model-based strategy) values (see Fig. 1c) from individual subjects (points). *rho* values are semi-partial Spearman correlation coefficients of per-subject information complexity versus main effect (a) or interaction (b) values while accounting for the other value; *p*-values are for $H_0$: *rho*=0. Plotted points represent residuals after accounting for interaction (a) and main effect (b) values. Lines are least-squares linear fits.

*Information complexity relates similarly to accuracy, optimality, and response times, irrespective of strategy type*

Thus, increased information complexity was associated with a stronger tendency to use either a model-free or a model-based strategy. Below we show that the two strategy types showed further similarities in terms of how they used these increases in information complexity. Specifically, we assessed the effectiveness (predictive accuracy), optimality (distance from optimal bound), and response times (mean log RT) of the strategy used by each subject with respect to its information complexity, model-free tendency (main effect term in Fig. 1c), and model-based tendency (interaction term in Fig. 1c).

The effectiveness of a given strategy tended to increase systematically as a function of its information complexity (Fig. 5a). This relationship was similar for both strategy types, such that predictive accuracy was slightly higher for more model-free and more model-based strategies (Fig. 5d,g). Thus, a tendency towards either pure strategy was associated with slight overall increases in performance.

The optimality of a given strategy tended to decrease systematically as a function of its information complexity (Fig. 5b). This relationship was similar for both strategy types, such that the distance was farther from the optimal bound for more model-free and more model-based strategies (Fig. 5e,h). Thus, despite the overall increases in accuracy, subjects using strategies with increasing information complexity made increasingly inefficient use of that information, in a manner that was similar for model-free and model-based strategies.

Mean log RT, measured for each subject across all trials, tended to increase systematically as a function of the information complexity of the subject's strategy (Fig. 5c). This relationship was similar for both strategy types, such that mean log RTs were higher for more model-free and more model-based strategies (Fig. 5f,i). The relationship with model-free tendencies was particularly striking, given that model-free strategies are commonly framed as automatic, habitual type responses, which should predict shorter RTs for subjects using more model-free strategies.

Thus, subjects with more complex strategies have improved predictive accuracy, but use their strategies in a less optimal manner and with longer RTs than subjects with less complex strategies. Moreover, these effects hold regardless of the source of information complexity, either

from associating past first responses with reward (more model-free strategies) or using other parts of the past feature space (more model-based strategies).

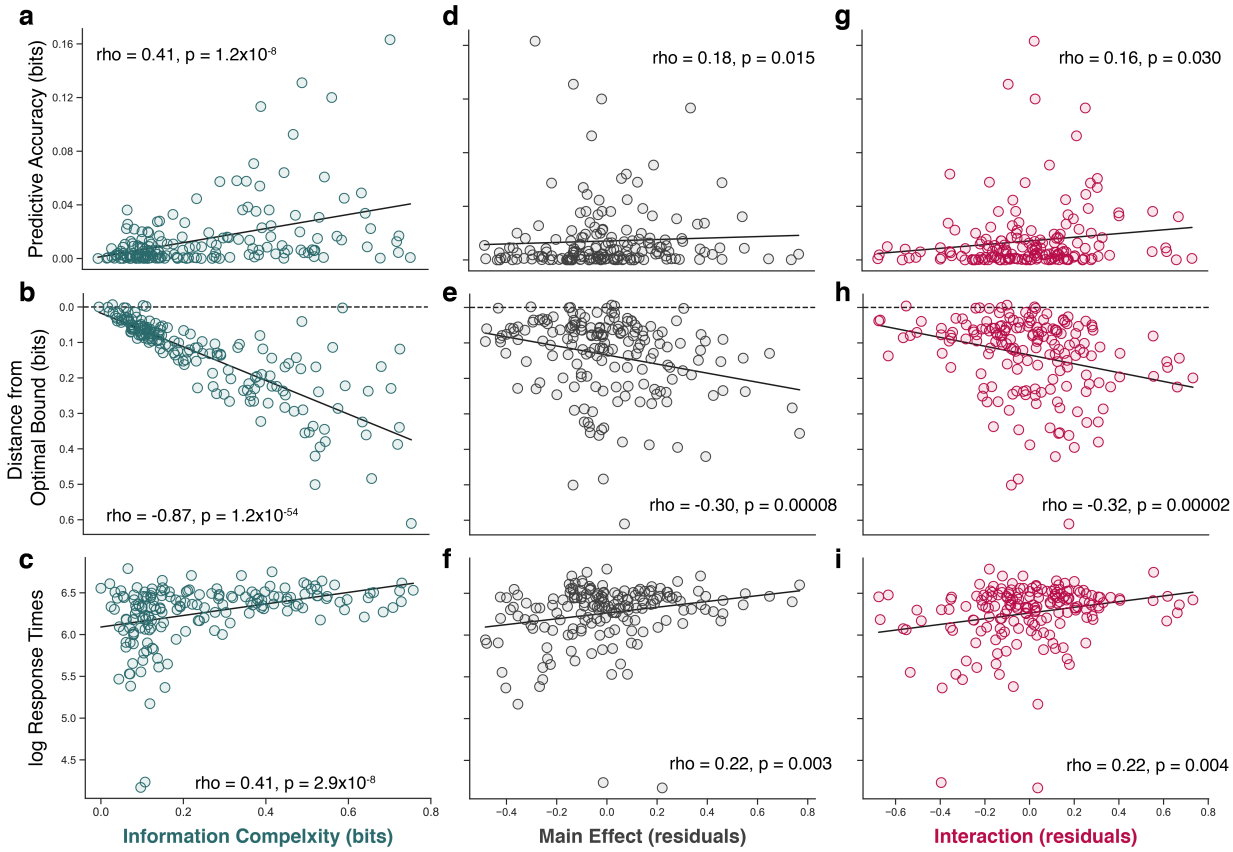


*Figure 5*. Comparison of subject information complexity and behavioral measures on predictive accuracy, optimality, and RT values. Subject with higher information complexity had higher predictive accuracy (a), but lower optimality (i.e., farther distance from the optimal bound; b) and slower RTs (c). These tendencies held irrespective of whether the information complexity was related to increased model-free or model-based behavior, as measured by main effects and interaction values respectively (d-i). Correlation values for main effect and interaction values indicate semi-partial Spearman correlation coefficients while accounting for the other value, respectively. Plotted main effect and interaction values in (d-i) correspond to residuals after accounting for interaction and main effect values, respectively. Lines are linear least-squares fits.

## Discussion

We measured differences in complexity between model-based and model-free learning strategies for a canonical learning task and report three main findings. First, model-based simulations required more operations to perform (higher computational complexity) but encoded less past information to guide future choices (lower information complexity) than model-free strategies. Second, human subject tendencies towards either model-free or model-based strategies were, unlike the simulations, both associated with higher information complexity, a measure of the flexibility with which different patterns of past observations were used to guide choices. Third, these increases in information complexity corresponded to higher predictive accuracy but less-optimal use of past information and longer RTs, irrespective of whether the source of increased information complexity was based on a model-free or model-based strategy.

Our results highlight the importance of distinguishing different forms of complexity, which covary under some but not all conditions. For example, increasing computational complexity by providing more computational resources can, in principle, support increased flexibility in how information is processed and therefore higher information complexity if these computations are used to expand the feature space over which inference is performed (Feldman & Crutchfield, 1998; Griffiths & Tenenbaum, 2003). However, as we showed, this relationship does not always hold: a standard model-free learning algorithm, while less computationally complex, is more information complex than a standard model-based learning algorithm.

We focused on the information complexity of the strategies used by human subjects, since we could measure this form of complexity directly from behavior. In contrast, we could not measure computational complexity from behavior without strong, and likely incorrect, assumptions about the specific algorithms subjects used to perform the task. Our simulations with commonly used learning models predicted that that model-free strategies should be more information complex than model-based strategies. However, we found that information complexity instead increased for strategies used by human subjects that were increasingly either model-free or model-based. The increase in information complexity for model-based strategies implies a strong discrepancy between the model-based algorithm we used in our simulations and the actual strategies used by the subjects. As has been highlighted recently, individual subjects likely use a variety of strategies that can appear similarly model-based in the two-step task, even if they differ substantially from the common model-based strategy used to fit their responses (Akam, Costa, &

Dayan, 2015; Silva & Hare, 2019). Moreover, there are likely substantial individual differences in the exact nature of these strategies, making it even more difficult to assess their computational complexity (Silva & Hare, 2019).

Our analyses of information complexity oppose the notion that model-free decision strategies are necessarily less complex, less deliberative, more automatic, more inflexible, and more habitual than model-based strategies used by human subjects performing the two-step task (Daw et al., 2011; Decker et al., 2016; Eppinger et al., 2013; Gläscher et al., 2010; Pauli et al., 2018). The increases in information complexity that we found for increasingly model-free or model-based strategies were accompanied by similar increases in overall accuracy, decreases in the degree to which past information was used optimally to guide behavior, and increases in RT for the two strategy types. Given that shorter RTs have been associated with habitual, automatic processing, such as for certain forms of statistical learning (Filipowicz, Anderson, & Danckert, 2014; Jabar, Filipowicz, & Anderson, 2017; Nissen & Bullemer, 1987; Robertson, 2007; Turk-Browne, Jungé, & Scholl, 2005), these results imply that the primary difference between model-free versus model-based strategies is not automaticity versus flexibility. Instead, both strategies can make flexible use of past task features to guide future behavior, but model-free strategies focus more on observed task features (e.g., responses, rewards, stimuli), whereas model-based strategies focus more on latent task features (e.g., transition structures).

Our results also highlight the usefulness of the information bottleneck for assessing performance optimality across a range of strategies (Tavoni, Balasubramanian, & Gold, 2019). The main strength of this approach over others is that it measures optimality with respect to the amount of information being encoded by that strategy and without requiring explicit knowledge of the underlying strategy (Gilad-Bachrach et al., 2003; Palmer et al., 2015; Tishby et al., 2000). Information-bottleneck analyses have been used previously to evaluate complexity-optimality tradeoffs in machine learning (Gilad-Bachrach et al., 2003; Tishby & Zaslavsky, 2015) and biological systems (Palmer et al., 2015). This kind of analysis is similar to notions of bounded rationality (Gigerenzer & Gaissmaier, 2011; Simon, 1955) and resource rational decision-making (Lieder & Griffiths, 2019; Tavoni, Doi, et al., 2019) that are becoming more prominent in assessing the rationality of human decision-making.

By applying the information bottleneck to simulated model-free and model-based learners, we showed that fully model-based simulations can be more computationally complex but encode

less information from past task features to achieve similar levels of predictive accuracy as model-free simulations. Under these conditions, the model-based strategy thus makes more optimal use of the information being encoded than the model-free strategy. In contrast, human subjects using more information-complex strategies used the increasing amounts of information they encoded less effectively (Fig. 5c), irrespective of whether they tended towards model-based or model-free strategies. This result extends our understanding of optimality for this task, which previously was defined in terms of average payouts but did not take into account the nature and amount of information used by subjects to achieve those payouts (Kool et al., 2016). Future work should aim to better understand relationships between information complexity and optimality, including how their balance is controlled by different individuals under different task conditions.

In summary, our results show that model-free and model-based learning strategies, often described as representing different ends of a continuum of information-processing flexibility, instead can be quite similar in terms of how much, how effectively, and how quickly they process information to perform a canonical learning task. These results imply that rather than distinguishing the flexibility of different learning processes, akin to the distinctions between automatic and deliberative or habitual and goal-directed processing, tasks such as the two-step task may instead distinguish between strategies that are equally complex but learn from different task features. A better understanding of these distinctions will help understand how and when these processes should be expected to vary across different healthy and psychiatric populations.

## Methods

### *Participants and behavioral task*

Human data were obtained from a publicly available dataset (Kool et al., 2016; https://github.com/wkool/tradeoffs). This data set consisted of 197 subjects, recruited on Amazon Mechanical Turk, who performed the two-step task that Kool and colleagues describe as the "Daw two-step task" (see Kool et al., 2016 for full task and subject details). All available data were used for our analyses without any additional exclusions.

### *Computational models*

We used the same computational models described by Kool and colleagues to simulate model-free and model-based agents (Kool et al., 2016). Fully model-free agents were simulated

using a SARSA(λ) temporal difference learning algorithm that updates Q-values ($Q_{MF}$) of stimulus-action pairs $(s_{i,j}, a_{i,j})$, where *i* indicates the step (first = 1, second = 2) and *j* denotes the state (only used for second step states given that there is only one first-step state). The Q-values at the first step are updated according to

$$Q_{MF}(s_1, a_1) = Q_{MF}(s_1, a_1) + \alpha\left[Q_{MF}\left(s_{2,j}, a_{2,j}\right) - Q_{MF}(s_1, a_1)\right] \quad (1)$$

where $\alpha$ corresponds to a fixed learning rate that determines the extent to which current values are modified by the reward prediction error, which at the first step is driven by the difference between the value of the action taken at the second step and the action taken at the first step $Q_{MF}\left(s_{2,j}, a_{2,j}\right) - Q_{MF}(s_1, a_1)$. When rewards are observed at the second step ($r_2$), the Q-value for the second step action is updated using the observed reward

$$Q_{MF}\left(s_{2,j}, a_{2,j}\right) = Q_{MF}\left(s_{2,j}, a_{2,j}\right) + \alpha\delta_2 \quad (2)$$

where

$$\delta_2 = r_2 - Q_{MF}(s_{2,j}, a_{2,j}) \quad (3)$$

The first level Q-values are then updated again as a function of the second step prediction error weighted by an eligibility trace decay parameter $\lambda$ such that when $\lambda = 0$, only the values of second step get updated

$$Q_{MF}\left(s_{1,}, a_1\right) = Q_{MF}(s_1, a_1) + \alpha\delta_2\lambda \quad (4)$$

Model-based simulations used the transition probability structure of the task to select their actions. First step model-based Q-values ($Q_{MB}$) are calculated using Bellman's equation, which uses the transition structure of the environment to determine expected values for each of the first step actions by deriving expectations over second step actions:

$$Q_{MB}(s_1, a_k) = p\left(s_{2,A} \middle| s_1, a_k\right) \max_{a_{2,A} \in \{a_{2,A,1}, a_{2,A,2}\}} Q_{MF}\left(s_{2,A}, a_{2,A}\right) \tag{5}$$

$$+ p(s_{2,B} | s_1, a_k) \max_{a_{2,B} \in \{a_{2,B,1}, a_{2,B,2}\}} Q_{MF}(s_{2,B}, a_{2,B})$$

which is performed for the *k* possible first step actions. Q-values at the second step are updated in the same manner as the model-free algorithm. These strategies, therefore, only differ in the computations they use to determine their first step action values.

To measure complexity of strategies mixing model-based and model-free action policies, mixed strategies were implemented with a $w$ strategy mixing coefficient, where $w = 1$ indicates a fully model-based decision strategy.

$$Q_{mix}(s_1, a_1) = wQ_{MB}(s_1, a_1) + (1 - w)Q_{MF}(s_1, a_1) \tag{6}$$

For all models, Q-values at each step are converted into action probabilities by applying a softmax function

$$p(a_i | s_i) = \frac{\exp\left[\beta Q_{mix}(s_i, a_i)\right]}{\sum_{a_i'} \exp\left[\beta Q_{mix}\left(s_i, a_i'\right)\right]} \tag{7}$$

where $\beta$ corresponds to an inverse temperature parameter that controls the randomness of the choice as a function of the Q-values (i.e., as $\beta \to 0$ action probabilities tends to become uniform, and as $\beta \to \infty$ the probability of choosing the action with the highest Q-value tends towards 1).

For our simulations, we performed 100 simulations of 10,000 trials for values of $w$ that varied between 0 and 1 in increments of 0.1. Information values for each value of $w$ were averaged across all 100 simulations. Apart from differences in the $w$ mixing parameter, all other parameters were kept constant in each simulation, consistent with the values used by Kool and

colleagues ($\alpha = 0.5$, $\beta = 5$, $\lambda = 0.5$). All of the trends reported in the manuscript hold for a wide range of parameter values, except when $\alpha \to 0$ (no learning occurs) or $\beta \to 0$ (behavior is random). Subject model fits, and behavioral metrics (main effect and interaction terms) we report were those fit and reported by Kool and colleagues (see Kool et al., 2016).

*Measure of information complexity and predictive accuracy*

We computed the information complexity and predictive accuracy of each subject and simulation using the information bottleneck method (Tishby et al., 2000). This method assumes that subjects and simulations are forming an internal belief or model $M$ from past task observations ($X_{past}$) to predict some aspect of the future ($X_{future}$). The amount of information $M$ encodes from $X_{past}$ is measured by their mutual information, i.e., $I_{past} = I(X_{past}; M)$, with high information corresponding to more information complex models (Filipowicz et al., 2019; Gilad-Bachrach et al., 2003; Glaze et al., 2018). Predictive accuracy is measured as the mutual information between the model and future observations, (i.e., $I_{future} = I(M; X_{future})$; Gilad-Bachrach et al., 2003; Palmer et al., 2015).

To compute $I_{past}$ we identified the observed and latent variables that could be used to perform the task. These included the first step response ($R1$; observed), second step transition ($S2$; observed), reward after second step response ($Rw$; observed) and the first step response that would maximize the chance of obtaining a reward ($R1^*$; latent). These four elements were combined on each trial to form past trial "features" ($F$), which could take 16 unique possible values. Subject and simulation information complexity was computed by measuring the mutual information between features from the previous trial $F_{-1}$ and first step responses on the current trial $R1_0$:

$$\text{Information Complexity} = I(F_{-1}; R1_0) = \sum_{f \in F_{-1}} \sum_{r \in R1_0} p(f,r) \log_2 \left[\frac{p(f,r)}{p(f)p(r)}\right] \quad (8)$$

Predictive accuracy was measured between subject first step responses ($R1_0$) and the aspect of the task they were attempting to predict (i.e., $R1_0^*$, the best action to take given the current task contingencies in order to maximize reward)

$$\text{Predictive Accuracy } = I(R1_0; R1_0^*) = \sum_{r \in R1_0} \sum_{r^* \in R1_0^*} p(r, r^*) \log_2 \left[ \frac{p(r,r^*)}{p(r)p(r^*)} \right] \quad (9)$$

Given the Markovian nature of both the process generating the stimuli and the processes guiding our simulations, past features included only elements of the previous trial and did not extend further in the past. Similar to previous applications of this method to experimental data, we also assumed that subjects treated the task in a Markovian manner, by including the latent variable $R1^*$ in the past feature as a proxy for the history of previously observed transitions (Filipowicz et al., 2019; Glaze et al., 2018). We also omitted the second-step response from the past feature vector, because the information provided by this past element does not inform the first-step responses in our simulations beyond the information provided by the combination of the second-step transition (*S2*) and the reward (*Rw*), and including this element in the past feature vector did not improve our ability to distinguish between model-based and model-free simulations. Both of these measures were implemented to help limit the dimensionality of the joint distribution between subject responses and past stimulus features $p(F, R1)$, because high dimensionality distributions are difficult to estimate precisely without the appropriate amount of data (i.e., the "curse of dimensionality"; Bellman, 1961).

*Measure of optimality*

To measure the optimality of subject and simulation behavior, we compared the predictive accuracy of each subject/simulation given their respective information complexity to the optimal information bottleneck bound (Gilad-Bachrach et al., 2003; Palmer et al., 2015; Tishby et al., 2000). We used this method to identify the extent to which information from the past ($X_{past}$) can be compressed into a belief or mental model ($M$) to maximally predict the future ($X_{future}$). The bound is estimated by finding the smallest compression of the past that also retains a particular level of predictive accuracy $I(M; X_{future})$. We do so by minimizing the functional

$$\mathcal{L}\left[p\left(M \mid X_{past}\right)\right] = I\left(X_{past}; M\right) - \beta I(M; X_{future}) \quad (10)$$

where $\mathcal{L}\left[p\left(M|X_{past}\right)\right]$ is the quantity to be minimized and $\beta$ is a Lagrange multiplier. By finding the minimum of eq. 10 for a range of values of $\beta$, we obtain an information bottleneck curve that

provides the best predictive accuracy ($I_{future}$) for a range of values of information complexity ($I_{past}$; Fig. 2b). When $\beta \to 0$, none of the past is retained and the compression $M$ has no predictive accuracy. Conversely, when $\beta \longrightarrow \infty$, all of the past is retained (i.e., no compression) and the curve saturates at the predictive information of the observations $I(X_{past}; X_{future})$ (Bialek et al., 2001).

We used the Blahut-Arimoto algorithm to compute the bound for a number of different values of $\beta$ (Tishby et al., 2000). The algorithm iterates through the following three equations, letting $n$ denote the iteration step

$$p_n(m|x_{past}) = \frac{p_n(m)}{\mathrm{Z}(X_{past}, \beta)} \exp\left[-\beta \sum_{x_{future}} p_n(x_{future}\,|x_{past}) \log_2 \frac{p_n(x_{future}|x_{past})}{p_n(x_{future}|m)}\right] \tag{11}$$

where $\mathrm{Z}(X_{past}, \beta)$ represents a normalizing factor,

$$p_{n+1}(x_{future}|m) = \frac{1}{p_n(m)} \sum_{x_{past}} p_n(x_{future}|x_{past}) p_n(m|x_{past}) p_n(x_{past}) \tag{12}$$

to ensure the Markovian relation $M \leftarrow X_{past} \longleftarrow X_{future}$, and

$$p_{n+1}(m) = \sum_{x_{past}} p_n(m|x_{past}) p_n(x_{past}) \tag{13}$$

Iterating through eq.11-13 until convergence provides a distribution $p(M|X_{past})$ for a given value of $\beta$ that minimizes information complexity $I(X_{past}; M)$ while maximizing predictive accuracy $I(M; X_{future})$ (see Tishby et al., 2000).

Given the Markovian nature of the stimuli used in this task, we used this algorithm to compute the information bottleneck between features from the previous trial $F_{-1}$ (which

correspond to $X_{past}$ ) and the optimal response on the current trial $R1_0^*$ (which corresponds to $X_{future}$ ) for different values of $\beta$. Our Python implementation of this algorithm, along with helper functions to apply it to empirical time-series data, is available at https://pypi.org/project/embo/.

**Acknowledgements**

The authors would like to thank Wouter Kool for giving permission to use the human subject data and for providing clear, complete, and easily reproducible simulation code. We would also like to thank Adrian Radillo for insightful comments on the manuscript and Songhan Zhang for help with simulations. Funded by NSF-NCS 1533623, R01 EB026945, and NIMH F32 MH117924. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author Contributions**

A.L.S.F., J.L., and E.P. adapted the information bottleneck measure to the two step task; J.L. performed the model simulations; A.L.S.F. and J.L. analyzed the human and simulation performance; E.P. and A.L.S.F. designed and implemented the software to empirically compute the information bottleneck bound; G.T. and A.L.S.F. performed the algorithmic complexity analyses; all authors interpreted the results and drafted and revised the manuscript.

**References**

Akam, T., Costa, R., & Dayan, P. (2015). Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. *PLoS Computational Biology*, *11*(12), 1–25. https://doi.org/10.1371/journal.pcbi.1004648

Bellman, R. (1961). *Adaptive Control processes: A guided tour.* Princeton, NJ, USA: Princeton University Press.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability , Complexity , and Learning. *Neural Computation*, *13*(11), 2409–2463.

Bossaerts, P., Yadav, N., & Murawski, C. (2019). Uncertainty and computational complexity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *374*(1766), 20180138. https://doi.org/10.1098/rstb.2018.0138

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based

influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychological Science*, *27*(6), 848–858. https://doi.org/10.1177/0956797616639301

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*(6), 1075–1081. https://doi.org/10.1016/j.conb.2012.08.003

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems.* Cambridge, MA, USA: MIT Press.

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*(7 DEC), 1–14. https://doi.org/10.3389/fnins.2013.00253

Feldman, D. P., & Crutchfield, J. P. (1998). Measures of statistical complexity: Why? *Physics Letters, Section A: General, Atomic and Solid State Physics*, *238*(4–5), 244–252. https://doi.org/10.1016/S0375-9601(97)00855-4

Filipowicz, A., Anderson, B., & Danckert, J. (2014). Learning what from where: Effects of spatial regularity on nonspatial sequence learning and updating. *Quarterly Journal of Experimental Psychology*, *67*(7). https://doi.org/10.1080/17470218.2013.867518

Filipowicz, A. L. S., Glaze, C. M., Kable, J. W., & Gold, J. I. (2019). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *BioRxiv*, (736140), 1–39. https://doi.org/10.1101/736140.

Filipowicz, Alex, Anderson, B., & Danckert, J. (2016). Adapting to change: The role of the right hemisphere in mental model building and updating. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(3), 201–218. https://doi.org/10.1037/cep0000078

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilad-Bachrach, R., Navot, A., & Tishby, N. (2003). An information theoretic tradeoff between

complexity and accuracy. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *2777*, 595–609. https://doi.org/10.1007/978-3-540-45167-9_43

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective and Behavioral Neuroscience*, *15*(3), 523–536. https://doi.org/10.3758/s13415-015-0347-6

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Glaze, C. M., Filipowicz, A. L. S., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213–224. https://doi.org/10.1038/s41562-018-0297-4

Griffiths, T. L., & Tenenbaum, J. B. (2003). Probability, algorithmic complexity, and subjective randomness. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25), 480–485.

Grünwald, P., & Rissanen, J. (2007). *The Minimum Description Length Principle*.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The Elements of Statistical Learning. *The Mathematical Intelligencer*, *27*(2), 83–85. https://doi.org/10.1198/jasa.2004.s339

Jabar, S. B., Filipowicz, A., & Anderson, B. (2017). Knowing where is different from knowing what: Distinct response time profiles and accuracy effects for target location, orientation, and color probability. *Attention, Perception, and Psychophysics*, *79*(8), 2338–2353. https://doi.org/10.3758/s13414-017-1412-8

Kim, D., Park, G. Y., O'Doherty, J. P., & Lee, S. W. (2018). Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *BioRxiv*, 1–34. https://doi.org/10.1101/393983

Kool, W., Gershman, S. J., & Cushman, F. A. (2018). Planning Complexity Registers as a Cost in Metacontrol. *Journal of Cognitive Neuroscience*, *30*(10), 1391–1404. https://doi.org/10.1162/jocn_a_01263

Kool, Wouter, Cushman, F. A., & Gershman, S. J. (2016). When Does Model-Based Control Pay Off? *PLoS Computational Biology*, *12*(8), 1–34. https://doi.org/10.1371/journal.pcbi.1005090

Kool, Wouter, Gershman, S. J., & Cushman, F. A. (2017). Cost-Benefit Arbitration Between

Multiple Reinforcement-Learning Systems. *Psychological Science*, *28*(9), 1321–1333. https://doi.org/10.1177/0956797617708288

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 1–85. https://doi.org/10.1017/S0140525X1900061XP

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, *97*(21), 11170–11175. https://doi.org/10.1073/pnas.170283897

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Nissen, M. J., & Bullemer, P. (1987). Attentional Requirements of Learning : Performance Measures Evidence from. *Cognitive Psychology*, *19*, 1–32.

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*(7 JUN), 1–10. https://doi.org/10.3389/fnins.2013.00105

Palmer, S. E., Marre, O., Berry, M. J., & Bialek, W. (2015). Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, *112*(22), 6908–6913. https://doi.org/10.1073/pnas.1506855112

Pauli, W. M., Cockburn, J., Pool, E. R., Pérez, O. D., & O'Doherty, J. P. (2018). Computational approaches to habits in a model-free world. *Current Opinion in Behavioral Sciences*, *20*, 104–109.

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior*, *94*, 80–96. https://doi.org/10.1016/j.geb.2015.09.003

Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning? *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(38), 10073–10075. https://doi.org/10.1523/JNEUROSCI.2747-07.2007

Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., … Huys, Q. J. M. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, *70*(2), 122–131. https://doi.org/10.1159/000362840

Silva, C. F. da, & Hare, T. A. (2019). Model-free or muddled models in the two-stage task?

*BioRxiv*, 682922. https://doi.org/10.1101/682922

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. Retrieved from http://www.jstor.org/stable/1884852

Sutton, R., & Barto, A. (1998). *Introduction to reinforcement learning*. Cambridge, MA: MIT Press.

Tavoni, G., Balasubramanian, V., & Gold, J. I. (2019). What is optimal in optimal inference? *Current Opinion in Behavioral Sciences*, *29*, 117–126. https://doi.org/10.1016/j.cobeha.2019.07.008

Tavoni, G., Doi, T., Pizzica, C., Balasubramanian, V., & Gold, J. I. (2019). The complexity dividend: when sophisticated inference matters. *BioRxiv*, 563346. https://doi.org/10.1101/563346

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science (New York, N.Y.)*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. *ArXiv Preprint Physics/0004057*, 1–16. https://doi.org/10.1108/eb040537

Tishby, N., & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. *2015 IEEE Information Theory Workshop, ITW 2015*, 1–5. https://doi.org/10.1109/ITW.2015.7133169

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., … Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352. https://doi.org/10.1038/mp.2014.44