Causal Model Comparison Shows That Human Representation Learning Is Not Bayesian

Andra Geana[1] and Yael Niv[1,2]
[1]*Psychology Department, Princeton University, Princeton, New Jersey 08540*
[2]*Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08540*
Correspondence: ageana@princeton.edu; yael@princeton.edu

How do we learn what features of our multidimensional environment are relevant in a given task? To study the computational process underlying this type of "representation learning," we propose a novel method of *causal model comparison*. Participants played a probabilistic learning task that required them to identify one relevant feature among several irrelevant ones. To compare between two models of this learning process, we ran each model alongside the participant during task performance, making predictions regarding the values underlying the participant's choices in real time. To test the validity of each model's predictions, we used the predicted values to try to perturb the participant's learning process: We crafted stimuli to either facilitate or hinder comparison between the most highly valued features. A model whose predictions coincide with the learned values in the participant's mind is expected to be effective in perturbing learning in this way, whereas a model whose predictions stray from the true learning process should not. Indeed, we show that in our task a reinforcement-learning model could help or hurt participants' learning, whereas a Bayesian ideal observer model could not. Beyond informing us about the notably suboptimal (but computationally more tractable) substrates of human representation learning, our manipulation suggests a sensitive method for model comparison, which allows us to change the course of people's learning in real time.

We live in a rich, complex environment, in which we are constantly bombarded with a wide variety of sensory input. Even an action as simple as walking down the street carries with it a large volume of low-quality information in the form of people we see, places we walk by, cars, colors, voices, noises, emotional content, etc. Intuitively, one would imagine that given sufficient resources, it is best to always represent every aspect of the environment so that any detail can potentially be acted upon. However, the "curse of dimensionality" (Bellman 1957) posits that task representations that involve unnecessary stimulus dimensions will not afford efficient learning and decision-making, where efficiency is measured in the number of examples needed to learn the task. In particular, an increase in the number of dimensions of the problem (in our case, the dimensions of the environment that the brain may represent) implies that the learner needs to collect exponentially larger quantities of data to learn to solve the problem. If we want learning to be feasible it is therefore both computationally optimal and a practical imperative to represent tasks with as compact a representation as possible.

What are the computational strategies that humans use to learn a representation for a given task? To address this question, we tested participants on a multidimensional trial-and-error choice task, in which only one dimension was relevant to predicting reward (Wilson and Niv 2012; Niv et al. 2015). To test the explanatory power of different models of learning dynamics, we developed a method that compares two models against each other in terms of their causal effects on behavior. Specifically, we used each model to manipulate participants' learning in real time, and asked which model was more effective in changing behavior. This is at the same time a novel, intuitive measure of how well a model captures participants' strategies, and it constitutes evidence that it is possible to use model predictions to impact learning in real time, by manipulating the stimuli that are presented to the participant.

It goes without saying that trial-and-error learning depends on what information is available to the learner. Indeed, work in machine learning and information theory has established how information in any given task might be optimally selected so as to maximally discriminate between competing hypotheses and accelerate learning (optimal experimental design; Sebastiani and Wynn 2000). Although human learning does not always mirror these optimal strategies, judicious choice of information has been shown to improve learning, for instance of category boundaries (Gureckis and Markant 2012) or speech motor learning (Knock et al. 2000). Moreover, the order in which information is presented is relevant to determining what is learned (Ritter 2007). We thus set forth to use our candidate models to manipulate the timing and availability of information in such a way as to aid or hinder participants' learning trajectory.

This kind of effort to manipulate learning, however, is heavily dependent on having a good model of how participants structure and update their representations of the environment. How to compare and select a "best"

Copyright © 2014 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/sqb.2014.79.024851

model for a complex cognitive process is not trivial (Cutting et al. 1992; Pitt et al. 2002): Models that fit the data better on some common goodness-of-fit measures may not fit better on other such measures; models that posit very different processes may perform similarly in terms of average fit (Townsend 1990; Rust et al. 1995); and a model that seems to describe behavior better might do so because of a more flexible function form or different numbers of parameters, and not necessarily because it better captures the underlying cognitive processes (Busemeyer and Wang 2000). We therefore developed an *interventional* method for model comparison.

We used our candidate models to predict in real time what hypotheses a participant might be testing and to design stimuli that will make it easier or harder to distinguish between the competing hypotheses. We reasoned that a model that does not capture the participants' beliefs about the available stimuli would not be effective at such a manipulation of learning. In contrast, a model whose predictions are well-matched to the cognitive processes underlying participants' behavior should allow us to manipulate learning in real time. In particular, while participants played our multidimensional probabilistic learning task, we inferred their value representations in real-time using either a Bayesian or a reinforcement-learning model. We used those inferred representations to present participants with information that would either help discriminate their competing hypotheses about which dimension of the stimuli is relevant for reward, or specifically hurt such a discrimination.

Despite much work suggesting that the human brain is Bayes-optimal (Körding and Wolpert 2004; Beierholm et al. 2009), and in line with our previous findings (Niv et al. 2015), our manipulation was only effective when we based it on predictions of the reinforcement-learning model. Our ability to manipulate the learning process both precisely and in real time consists of a proof of concept for the new proposed model-testing tool, and is a step in the right direction in terms of development of individualized tools to improve learning in general.

METHODS

Participants

Twenty-five participants (16 females) recruited from the Princeton University undergraduate community gave informed consent and were compensated $12 an hour plus a performance bonus of up to $5 depending on their final score in the task. The average pay was $15. Study materials and procedures were approved by the Princeton University Institutional Review Board.

Task

Participants played a probabilistic learning task. Each trial involved choosing one of three compound stimuli displayed on the screen (see Fig. 1A). Each stimulus was comprised of three features defined on three dimensions: a color (red, yellow, or green), a shape (triangle, square, or circle), and a texture (dots, plaid, or waves). No two stimuli could share the same feature (i.e., there was only one red stimulus, one triangle, etc., per trial). Choosing a stimulus resulted in immediate feedback in the form of either one or zero points. Participants were instructed to try to obtain as many points as possible.

The task was designed so that, of the three dimensions of a stimulus, only one dimension was relevant to determining reward. Within that dimension, one feature was the target feature—choosing the stimulus that contained this feature led to one point 75% of the time and zero points 25% of the time. Choosing any other stimulus resulted in one point only 25% of the time. To maximize their score, participants therefore had to aggregate over previous choices and outcomes to learn which feature is the target feature. Participants were explicitly instructed about these aspects of the task structure in advance.

The task consisted of 52 "games." Participants were informed that the target feature would not change within a game, but would change between games. Ends of games were explicitly signaled on-screen. The first 12 games (referred to as the baseline phase, described below) included 30 trials each, whereas the remaining games (the manipulation phase) consisted of 36 trials each, for a total of 1800 trials. After each game, participants were asked how difficult they found the game (on a scale from very easy [1] to very difficult [9]; Fig. 1B) and to identify the target feature in that game. They could select any of the nine features, as well as an "I do not know" option. If they did select a feature, they were also asked to rate how confident they were about their choice.

After the baseline phase, participants took a 1-min break, during which we used their baseline phase data to fit the free parameters of the two candidate models we would later test in the manipulation phase. The remaining 40 games of the task comprised of the manipulation phase, in which we manipulated stimuli according to predictions from each model, to either help or hurt participants' learning (see below).

Modeling

The two models we compared represent two different ways of thinking about human representation learning. The first is a Bayesian model that assumes statistically optimal updating of the posterior probabilities of each feature being the target feature, and the second model uses reinforcement-learning principles to update values via trial and error. Both models compute the value of a compound stimulus by estimating values of individual features and combining them: The Bayesian model estimates, for each feature, the posterior probability that it is the target feature, whereas the reinforcement-learning (RL) model learns the values of each feature based on prediction errors. In both models, current values of stimuli depend on the history of choices and rewards.

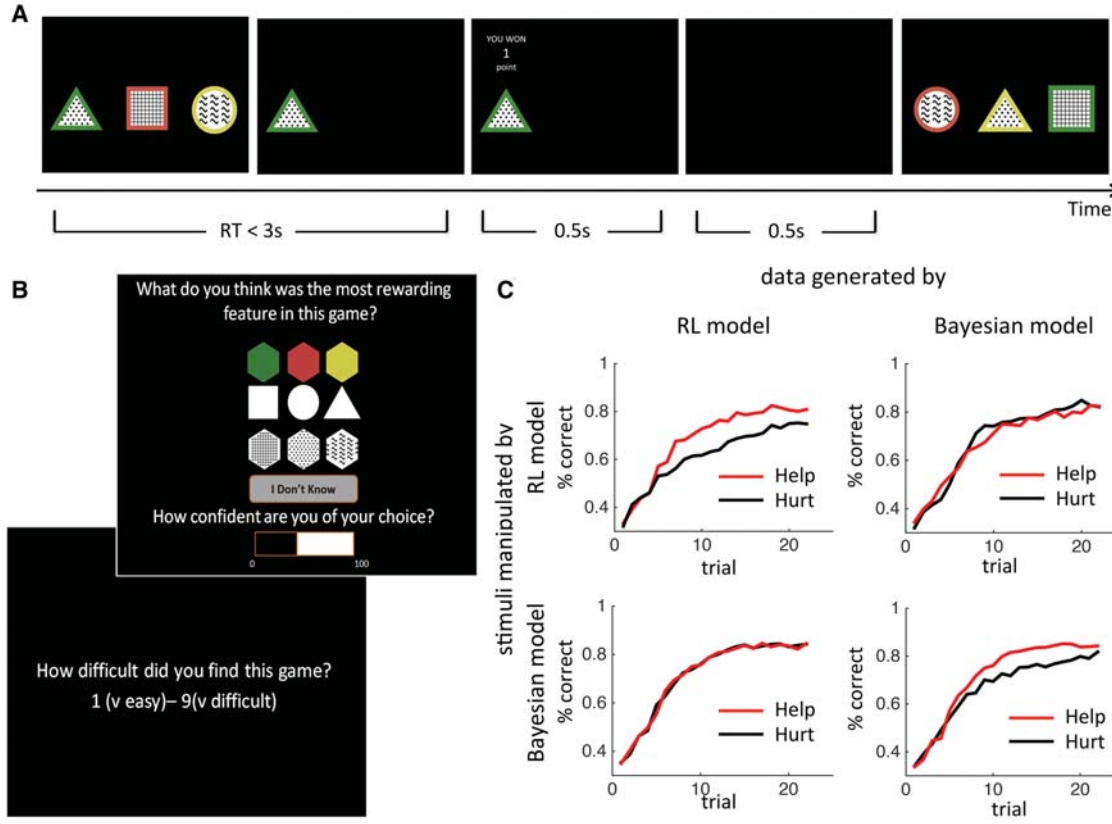

Figure 1. The dimensions task. (*A*) Example trial—stimulus presentation, choice, positive feedback, new stimulus presentation. (*B*) Query screens—difficulty ratings, identifying correct feature, confidence ratings. (*C*) Simulations of model-based manipulation. The manipulation was effective ("Help" improves performance as compared with "Hurt") only when stimuli were manipulated according to predictions from the same model that generated the choice data (*top left*, *bottom right* panels). The task was designed so that, of the three dimensions of a stimulus, only one dimension was relevant to determining reward. Within that dimension, one feature was the target feature—choosing the stimulus that contained this feature led to one point 75% of the time and zero points 25% of the time. Choosing any other stimulus resulted in one point only 25% of the time. To maximize their score, participants therefore had to aggregate over previous choices and outcomes to learn which feature is the target feature. Participants were explicitly instructed about these aspects of the task structure in advance.

Bayesian model. The Bayesian model tracks the posterior probability that each feature f is the target feature f^*. At the end of each trial, the posterior is updated by combining the prior (i.e., the posterior from the previous trial) and the likelihood of the observed data if f were the target feature. The prior depends on the history of choices C and rewards R, $D_{1,t-1} = \{C_{1:t-1}; R_{1:t-1}\}$, from the beginning of the game and up until the current trial (not inclusive). The likelihood depends on the reward probabilities imposed by the experimenter; for instance, the likelihood of a win if the chosen stimulus contains the target feature is 0.75.

At the beginning of the game, the prior is initialized at $1/9$ (all features are equally likely to be the target feature). After each trial, the posterior is updated according to

$$P(f = f^*|D_{1:t}) \propto P(R_t|f = f^*, C_t)P(f = f^*|D_{1:t-1}). \quad (1)$$

The value of a stimulus S is then calculated as the probability of obtaining a 1 point reward for choosing that stimulus on the current trial t,

$$\begin{aligned} V(S) &= P(R = 1|S, D_{1,t-1}) \\ &= \sum\nolimits_{f \in S} [P(R = 1|f = f^*)P(f = f^*|D_{1,t-1})] \\ &\quad + P(R = 1|f^* \notin S) \\ &\quad \times \left(1 - \sum\nolimits_{f \in S} P(f = f^*|D_{1,t-1})\right), \end{aligned} \quad (2)$$

where $P(R = 1|f = f^*) = 0.75$ for all features contained in S, and $P(R = 1|f^* \notin S) = 0.25$. The model can be considered an ideal observer because it maintains a full probability distribution over the identity of f^* and updates this distribution in a statistically optimal way.

To afford this model, the same temporal locality as the reinforcement model (described below), we also allowed some degree of decay for all feature posteriors toward a uniform value of 1/9,

$$\tilde{P}(f = f^*) = (1 - d)P(f = f^*) + d*1/9, \quad (3)$$

and used $\tilde{P}$ instead of P in Equations 1 and 2 above. Although suboptimal, the decay component has been shown to significantly improve the models' fit to behavioral choices in our task (Niv et al. 2015).

Finally, we assumed that the probability of choosing stimulus S_i on each trial is proportional to the value of the stimulus, according to the softmax probability choice function:

$$P(\text{choose } S_i) = \frac{e^{\beta V(S_i)}}{\sum_{j=1}^{3} e^{\beta V(S_j)}}. \quad (4)$$

The positive-valued inverse-temperature parameter β sets the level of noise in the decision process, with large β resulting in near-deterministic choice of the highest value option, whereas small β results in high decision noise and more random decisions. In all, this model has two free parameters, β and the decay rate d.

Reinforcement-learning model. This model takes advantage of the fact that, in our task, features determine reward independently, and uses reinforcement learning to learn values for each of the nine features. The values of all features were initialized at zero at the beginning of each game; on each trial, values of the three features of the stimulus that was chosen were then updated according to

$$V_t(f) = V_{t-1}(f) + \eta(R_t - V_{t-1}(S_{\text{chosen}})), \quad \forall f \in S_{\text{chosen}}, \quad (5)$$

where η represents the learning rate, and $[R_t - V_{t-1}(S_{\text{chosen}})]$ is a prediction error—the discrepancy between the actual reward on the current trial and the reward that was expected based on choosing this stimulus.

Based on our previous findings (Niv et al. 2015), we also included in this model a decay of the values of the unchosen features to zero:

$$V_t(f) = (1 - d)V_{t-1}(f), \quad \forall f \notin S_{\text{chosen}}, \quad (6)$$

with d a free parameter determining the decay rate. As in the Bayesian model, action probabilities were determined by a softmax probability function on stimulus values (Equation 4). The RL model thus had three free parameters, the learning rate η, the softmax inverse temperature β, and the decay rate d.

Model fitting. At the end of the baseline phase (described above), participants were given a 1-min break while the computer fit both models to their data. Free parameters of both models were set for each participant separately, and were selected so as to maximize the likelihood of the data from the baseline phase (12 games and a total of 360 trials), by using the Matlab routine fmincon and fitting the data five times, initializing at different random starting points. Parameter values from the run that obtained the best likelihood for the data were then used to fully specify the model for the manipulation phase of the experiment, in which we used each model to track feature values in real time.

Real-Time Manipulation

In the remaining 40 games of our task, we aimed to help or hurt learning by manipulating stimulus presentation. Specifically, we manipulated how the available features (colors, shapes, and textures) were combined into three different stimuli, so as to either make available or obscure information about which feature is more likely to be the target feature. This manipulation relies on the idea—common to both models we tested—that while playing the task, participants update values for each feature, with the goal of ultimately learning which is the most rewarding feature. These feature values carry predictions regarding the reward associated with each feature and thus can be seen as "hypotheses" as to which is the target feature.

For every manipulated trial, to help learning, the highest-valued feature in each dimension was selected, and each of the three stimuli presented to the participant was designed to include only one of these three highest-valued features. This manipulation facilitates hypothesis testing, allowing participants to test one high-value feature independently of the other two. Conversely, to hurt learning, the three highest-valued features (one in each dimension) were combined into a single stimulus, thus preventing the participant from assigning credit for the feedback to just one of the three competing features. Both types of manipulation can potentially impact the rate of learning in the game, but only as long as the inferred values are close to the participants' actual values (Fig. 1C).

In each manipulated game, we manipulated every other trial from the fourth to the thirtieth trial for a total of 14 manipulated trials, using one of the two models and either helping or hurting learning throughout the game. Because our manipulation affects not only learning, but the likelihood to make the correct choice on the current trial, to measure learning we analyzed choices only in nonmanipulated (neutral) trials, in which stimuli were constructed so as to specifically not include all three highest-valued features in the same stimulus, nor separate them into three different stimuli. (Therefore, in nonmanipulated trials, one stimulus always contained exactly two of the highest-valued features—and these trials did not overlap with either the helping or the hurting manipulation.)

Each of the 40 games in the manipulation phase consisted of 36 trials, with 14 manipulated and 22 neutral trials. The last six trials in each game were not manipu-

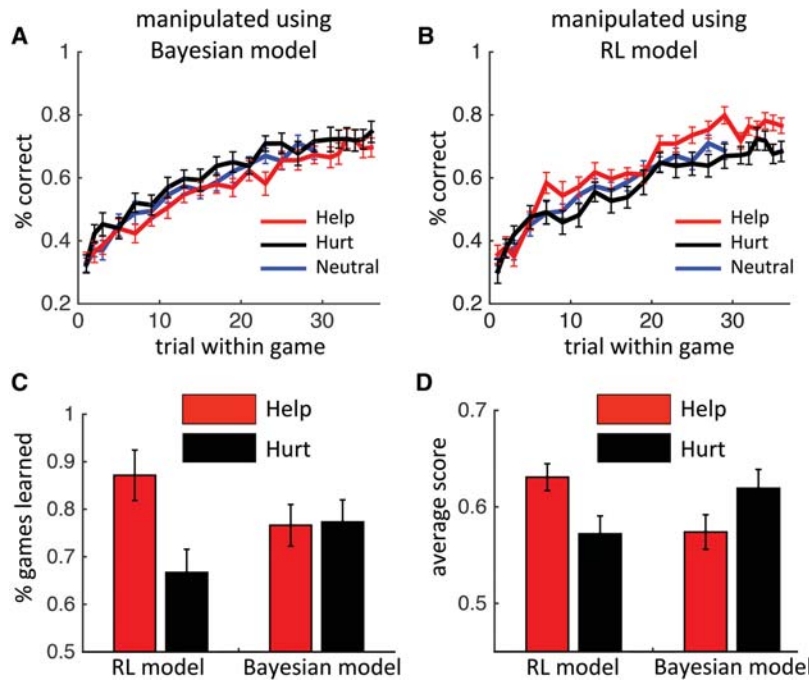

Figure 2. Model-based manipulation of stimuli affects learning only when using predictions from the RL model, not from the Bayesian model. (*A*) Learning curves for Help (red) and Hurt (black) conditions overlap when the manipulation used predictions from Bayesian model. (*B*) When the manipulation used predictions from the RL mode, learning curves for the Help condition show significantly better learning at the end of the game as compared with the Hurt condition. (Blue line) Data from the baseline phase. Similar effects of the manipulation are also seen in *C*, the overall number of learned games, and *D*, the average score across games. Error bars: S.E.M.

lated so as to allow measurement of steady-state learning at the end of the game.

RESULTS

To understand the dynamic process of learning a compact and sufficient task representation in a multidimensional environment, we tested human participants on a probabilistic three-dimensional choice task. While they played the task, we used the real-time trial-by-trial predictions from two competing models to manipulate the presented stimuli so as to help or hurt learning. Success of the causal manipulation would attest to congruence between the model and participants' learning strategies.

As shown in Figure 2, the manipulation had a significant effect on learning only in those games in which we manipulated stimuli using the RL model (Fig. 2A), and did not alter learning in games in which we manipulated stimuli using the Bayesian model (Fig. 2B). For RL-manipulated games, average performance on the last six trials of games (all nonmanipulated) in the Help condition (red line) was significantly better than in the Hurt condition (black line). A two-way Performance × Manipulation repeated-measures ANOVA yielded a significant interaction ($F_{(1,48)} = 5.22$, $P = 0.02$) with no main effects, and the average performance at the end of the game was significantly higher in the Help than the Hurt condition for RL-manipulated games (one-sided paired *t*-test, $t_{(24)} = 2.25$, $P = 0.01$), but did not differ between conditions for the Bayesian-manipulated games ($t_{(24)} = -0.96$, $P = 0.82$).

Participants did not learn all the games, that is, for some games, they could not identify the correct feature when probed at the end of the game (Fig. 1B). The real-time manipulation had an impact on the total number of games learned (Fig. 2C): Here too a two-way Model × Manipulation repeated-measures ANOVA showed a significant interaction ($F_{(1,48)} = 7.23$, $P = 0.009$) with no main effects. When the RL model was used to manipulate stimuli, the number of learned games differed significantly depending on whether the game was designed to help or hurt learning ($F_{(1,24)} = 12.64$, $P < 0.01$). On average, the Help condition resulted in an ~30% increase in the number of learned games. Conversely, when the Bayesian model was used to manipulate stimuli, there was no difference between the help and hurt conditions ($F_{(1,24)} = 0.77$, $P = 0.73$).

A similar pattern was observed in the average score per game (Fig. 2D). For games manipulated using the RL model, the average score (number of correct choices) was higher in the Help condition than in the Hurt condition (paired *t*-test; $t_{(24)} = 2.72$, $P = 0.011$). In contrast, scores for the Help and Hurt conditions were similar when the Bayesian model was used to manipulate the stimuli ($t_{(24)} = -1.91$, $P = 0.07$).

The impact of the stimulus manipulation was also evident in participants' difficulty (Fig. 3A) and confidence (Fig. 3B) ratings. For RL-manipulated games, participants rated "help" games as easier compared with the "hurt"

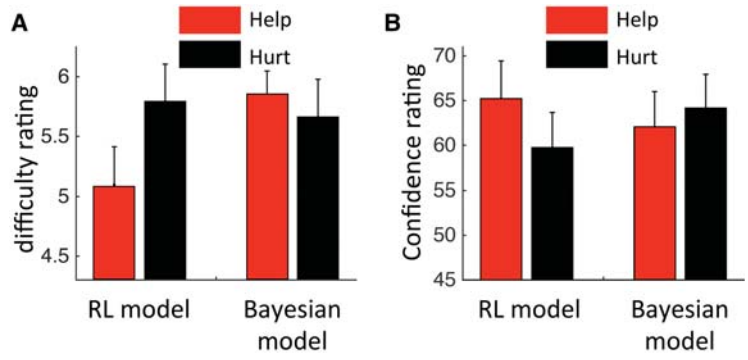

Figure 3. Difficulty (*A*) and confidence (*B*) ratings reflect the fact that the manipulation was only effective in RL-manipulated games, not in games manipulated using the Bayesian model. Error bars: S.E.M. Three subjects were excluded from this analysis, two because they consistently did not rate difficulty/confidence, and one participant who reported reversing the difficulty scale for most of the experiment.

games ($t_{(21)} = -2.23, P = 0.03$, paired *t*-test). This effect was not present in the games manipulated using the Bayes model ($t_{(21)} = 0.734$, $P = 0.47$, paired *t*-test). A similar pattern was seen in the confidence ratings (Fig. 3B), where confidence in RL-manipulated "help" games was rated as significantly higher than confidence in the "hurt" games ($t_{(21)} = 2.57, P = 0.01$), but ratings in games manipulated using the Bayesian model were not significantly different across conditions ($t_{(21)} = -1.26, P = 0.22$).

DISCUSSION

Using a multidimensional choice task, we investigated the computational strategies by which humans determine what dimensions of the environment are relevant for obtaining reward, and which can be safely ignored. The assumption underlying our work is that naturalistic tasks require such a representation learning process: In any given scenario, only a subset of information is relevant to the task at hand, and, moreover, the particular environmental dimensions that are relevant to one task are not necessarily relevant for another. For instance, the color of cars is irrelevant for crossing the street, but relevant for hailing a taxi, and the identity of the shops across the street is irrelevant to both tasks (but of course not for the task of navigating to the coffee shop).

To compare between qualitatively different accounts of how humans may learn what dimension of the environment is relevant for the current task, we showcased a novel method that compares two learning models by attempting to use each model in a causal, real-time manipulation of participants' learning. That is, we used each model to predict what hypotheses participants might be testing at each point in time, and manipulated stimuli to either help or hinder comparison of these hypotheses. This model-based manipulation can affect learning only to the extent that a model indeed captures participants' underlying learning process. We found that when stimuli were manipulated based on a RL model, games designed to help learning resulted in faster and more complete learning than games designed to hurt the learning process. In contrast, manipulating games using a Bayesian model had no significant effect on learning. Our method thus provides a stringent measure of how well each model captures people's strategies, and at the same time, our results provide evidence that it is indeed possible to impact representation learning in real time, by manipulating the stimuli that people have access to.

Our method is related to the framework of "optimal experimental design" in which experiments are designed so as to optimally elicit information about the process under investigation (Sebastiani and Wynn 2000; Atkinson et al. 2007). Normative statistical principles from Bayesian inference can, in some cases, be used to select an experimental design that will best resolve the details of participants' underlying cognitive processes (e.g., set the free parameters of a model of the process under scrutiny; Rafferty et al. 2012). One way to optimize our manipulation would be to choose, on each trial, the stimulus configuration that would allow participants to glean the maximum (or minimum) amount of information regarding the identity of the target stimulus, assuming participants' underlying cognitive processes matched each of the candidate models. This manipulation, although more normative than the one we designed, is more computationally intensive and, importantly, relies on further assumptions regarding the optimality of participants' actions. In particular, if participants are not selecting stimuli in an effort to maximize information (e.g., because they are also maximizing reward), this manipulation may not be more effective than ours. It is due to this interaction between information acquisition and reward acquisition that we assessed performance only in nonmanipulated trials—our "help" manipulation, although affording better information, made it more difficult to obtain reward on manipulated trials than did our "hurt" manipulation.

Rather than assume that the highest-valued features correspond to the hypotheses that the participant is comparing, another alternative is to infer these hypotheses using Bayesian inference. We have previously used such a method in association with the dimensions task and shown that recent choice history can effectively identify tested hypotheses (Wilson and Niv 2012). However, in

that work, inference was only optimal if we assume that participants test one hypothesis at a time—an assumption that is incompatible with our current results. If participants were indeed focusing only on one hypothesis (feature) at any point in time, then neither of our manipulations would have affected learning.

Our findings join others (Eckstein et al. 2004) in suggesting that human learning is not always Bayes-optimal and, in particular, that humans do not solve the difficult task of representation learning in a Bayes-optimal way (see also Wilson and Niv 2012; Niv et al. 2015). These findings stand in contrast to multiple demonstrations of Bayes-optimality (Doya 2007) in perceptual decision-making (Gold and Shadlen 2002; Knill and Pouget 2004), motor control (Trommershäuser et al. 2003, 2005), multimodal integration (Körding and Wolpert 2004), reasoning (Oaksford and Chater 1994), and even setting metalearning parameters for reinforcement learning (Behrens et al. 2007; Yu 2007). However, whereas Bayesian inference may be computationally feasible (and indeed, simple) in scenarios that can be reduced to a several-alternative forced-choice decision (Gold and Shadlen 2002) or a choice between lotteries (Wu et al. 2009), representation learning in natural environments places much heavier computational demands on the learning system. In particular, the dimensionality of the environment is essentially unbounded (given that dimensions such as previous actions and events can be, and frequently are, relevant for task performance), and whereas feedback is often available for one's actions, the environment does not provide any supervision regarding one's representations.

To overcome these difficulties, our results suggested that humans might use a suboptimal but computationally more tractable strategy based on reinforcement learning. However, we note that we only compared two very different models, partly as a proof-of-concept for our novel model-comparison method. It is entirely possible—in fact, likely—that the feature-level RL model that we suggested also falls short of fully capturing participants' learning strategies. Future applications of this method will hopefully delineate more precisely the workings of representation learning in the human brain.

Finally, our "dimensions task" lends itself easily to manipulation of presented information. Work on "active learning" (Cohn et al. 1996) in categorization and perceptual estimation tasks has used a related manipulation, effectively allowing participants to design their experiment optimally (Kruschke 2008; Castro et al. 2009; Gureckis and Markant 2009; Juni et al. 2011; Markant and Gureckis 2014). Some adjustments will likely be needed to apply this model-comparison method to other task structures, though we are optimistic as to the method's wider applicability (Nelson et al. 2010).

In sum, we have described a real-time manipulation of information presented to participants, and have suggested that basing this manipulation on predictions of different models can allow for a new, sensitive and *causal* means of model comparison. Using this method and a RL model, we have shown that human representation learning can be improved or hampered. Beyond the implications for effective, individual-difference-sensitive model selection, such "access" to participants' mental strategies suggests exciting applications, particularly in the domain of education and tailoring the flow of information toward individual learning.

ACKNOWLEDGMENTS

This work was funded in part by the National Institutes of Health grant R01MH098861 and Army Research Office award W911NF-14-1-0101. The information herein reflects the opinion of the authors and does not necessarily reflect the opinion or policy of the federal government.

REFERENCES

Atkinson A, Doney A, Tobias R. 2007. *Optimum experimental designs, with SAS*. (ed. Atkinson RJCA, Hand DJ, Pierce DA, Schervish MJ, Titterington DM). Oxford University Press, Oxford.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF. 2007. Learning the value of information in an uncertain world. *Nat Neurosci* **10:** 1214–1221.

Beierholm UR, Quartz SR, Shams L. 2009. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *J Vis* **9:** 23.

Bellman R. 1957. *Dynamic programming*. Princeton University Press, Princeton.

Busemeyer JR, Wang YM. 2000. Model comparisons and model selections based on generalization criterion methodology. *J Math Psychol* **44:** 171–189.

Castro RM, Kalish C, Nowak R, Qian R, Rogers T, Zhu X. 2009. Human active learning. In *Advances in neural information processing systems*, pp. 241–248.

Cohn DA, Ghahramani Z, Jordan MI. 1996. Active learning with statistical models. *J Artif Intell Res* **4:** 129–145.

Cutting JE, Bruno N, Brady NP, Moore C. 1992. Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *J Exp Psychol Gen* **121:** 364.

Doya K. (Ed.). 2007. *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press, Cambridge, MA.

Eckstein MP, Abbey CK, Pham BT, Shimozaki SS. 2004. Perceptual learning through optimization of attentional weighting: Human versus optimal Bayesian learner. *J Vis* **4:** 3.

Gold JI, Shadlen MN. 2002. Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36:** 299–308.

Gureckis T, Markant D. 2009. Active learning strategies in a spatial concept learning game. In *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*, pp. 3145–3150.

Gureckis TM, Markant DB. 2012. Self-directed learning: A cognitive and computational perspective. *Perspect Psychol Sci* **7:** 464–481.

Juni MZ, Gureckis TM, Maloney LT. 2011. Don't stop 'til you get enough: Adaptive information sampling in a visuomotor estimation task. In *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society* (ed. Carlson L, Hölscher C, Shipley T). Cognitive Science Society, Austin, TX.

Knill DC, Pouget A. 2004. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci* **27:** 712–719.

Knock TR, Ballard KJ, Robin DA, Schmidt RA. 2000. Influence of order of stimulus presentation on speech motor learning: A principled approach to treatment for apraxia of speech. *Aphasiology* **14:** 653–668.

Körding KP, Wolpert DM. 2004. Bayesian integration in sensorimotor learning. *Nature* **427:** 244–247.

Kruschke JK. 2008. Bayesian approaches to associative learning: From passive to active learning. *Learn Behav* **36:** 210–226.

Markant DB, Gureckis TM. 2014. Is it better to select or to receive? Learning via active and passive hypothesis testing. *J Exp Psychol Gen* **143:** 94.

Nelson JD, McKenzie CR, Cottrell GW, Sejnowski TJ. 2010. Experience matters: Information acquisition optimizes probability gain. *Psychol Sci* **21:** 960–969.

Niv Y, Daniel R, Geana A, Gershman S, Leong YC, Radulescu A, Wilson RC. 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci* **35:** in press.

Oaksford M, Chater N. 1994. A rational analysis of the selection task as optimal data selection. *Psychol Rev* **101:** 608.

Pitt MA, Myung IJ, Zhang S. 2002. Toward a method of selecting among computational models of cognition. *Psychol Rev* **109:** 472.

Rafferty AN, Zaharia M, Griffiths TL. 2012. Optimally designing games for cognitive science research. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pp. 893–898.

Ritter FE. 2007. *In order to learn: How the sequence of topics influences learning*. Oxford University Press, New York.

Rust RT, Simester D, Brodie RJ, Nilikant V. 1995. Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Manag Sci* **41:** 322–333.

Sebastiani P, Wynn HP. 2000. Maximum entropy sampling and optimal Bayesian experimental design. *J R Stat Soc Series B Stat Methodol* **62:** 145–157.

Townsend JT. 1990. Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol Sci* **1:** 46–54.

Trommershäuser J, Maloney LT, Landy MS. 2003. Statistical decision theory and the selection of rapid, goal-directed movements. *J Opt Soc Am A Opt Image Sci Vis* **20:** 1419–1433.

Trommershäuser J, Gepshtein S, Maloney LT, Landy MS, Banks MS. 2005. Optimal compensation for changes in task-relevant movement variability. *J Neurosci* **25:** 7169–7178.

Wilson R, Niv Y. 2012. Inferring relevance in a changing world. *Front Hum Neurosci* **5:** 189.

Wu SW, Delgado MR, Maloney LT. 2009. Economic decision-making compared with an equivalent motor task. *Proc Natl Acad Sci* **106:** 6088–6093.

Yu AJ. 2007. Adaptive behavior: Humans act as Bayesian learners. *Curr Biol* **17:** R977–R980.