Author's personal copy

Cognitive Psychology 58 (2009) 49–67

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Detecting and predicting changes

Scott D. Brown [a,*], Mark Steyvers [b]

[a] School of Psychology, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia
[b] University of California, Irvine, USA

ARTICLE INFO

Article history:
Accepted 9 May 2008
Available online 1 November 2008

Keywords:
Random sequences
Change detection
Particle filters
Gambler's Fallacy
Hot hand

ABSTRACT

When required to predict sequential events, such as random coin tosses or basketball free throws, people reliably use inappropriate strategies, such as inferring temporal structure when none is present. We investigate the ability of observers to predict sequential events in dynamically changing environments, where there is an opportunity to detect true temporal structure. In two experiments we demonstrate that participants often make correct statistical decisions when asked to infer the hidden state of the data generating process. However, when asked to make predictions about future outcomes, accuracy decreased even though normatively correct responses in the two tasks were identical. A particle filter model accounts for all data, describing performance in terms of a plausible psychological process. By varying the number of particles, and the prior belief about the probability of a change occurring in the data generating process, we were able to model most of the observed individual differences.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Many real-world environments involve complex changes over time, where behavior that was previously adaptive becomes maladaptive. These dynamic environments require a rapid response to change. For example, stock analysts need to quickly detect changes in the market in order to adjust investment strategies, and coaches need to track changes in a player's performance in order to adjust team strategy. Reliable change detection requires accurate interpretation of sequential dependencies in the observed data. Research on decision making in probabilistic environments has often called into

* Corresponding author.
E-mail address: scott.brown@newcastle.edu.au (S.D. Brown).
URL: http://science-it.newcastle.edu.au/~sdb231/ (S.D. Brown).

0010-0285/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2008.09.002

Author's personal copy

question our ability to correctly interpret sequential events. When required to predict random events, such as coin tosses, people reliably use inappropriate strategies, such as the famous "Gambler's Fallacy" identified by Kahneman and Tversky (1973, 1979)—if a fair coin is tossed several times, people often believe that a tail becomes more likely after a long run of heads (see also Burns & Corpus, 2004; Sundali & Croson, 2006). The error of reasoning that underlies the Gambler's Fallacy is the perception of probabilistic regularities in a sequence where no such regularities are present, because the sequence is in fact truly random. Such perceptions might arise because real-world environments rarely produce truly random sequences—often there really is statistical information in the sequence of events. This suggests that the Gambler's Fallacy could simply be the result of people projecting their experience of real-world environments onto laboratory tasks.

Related work in dynamical systems research using response time tasks paints a complementary picture. When the optimal strategy in a task is to provide a series of independent and identically distributed responses, people often perform sub-optimally. Long-range autocorrelations have been observed, where responses depend on earlier events that occurred quite a long time previously (e.g., Gilden, 2001; Thornton & Gilden, 2005; Van Orden, Holden, & Turvey, 2003, 2005), although not all authors agree on the meaning of the phenomena (e.g., Farrell, Wagenmakers, & Ratcliff, 2006; Wagenmakers, Farrell, & Ratcliff, 2004, 2005). The same criticism applies to dynamical systems research as to the Gambler's Fallacy—tasks requiring long sequences of stationary and conditionally random responses have questionable ecological validity.

Even with real-world environments, people often observe statistical regularities where no such regularities might be present (e.g., Albright, 1993; Gilovich, Vallone, & Tversky, 1985). For example, when a basketball player makes several successes in a row, observers readily conclude that the player's underlying skill level has temporarily increased; that the player has a "hot hand". Observers make these conclusions even when the data are more consistent with random fluctuations than with underlying changes in skill level. A problem with the hot hand phenomenon is the statistical interpretation of the results. The uncontrolled nature of batting averages and basketball successes make the true state of the underlying process impossible to know. Even after detailed statistical analyses of data from many games, statisticians are still unsure whether a "hot hand" phenomenon actually exists in the data (Adams, 1992; Chatterjee, Yilmaz, Habibullah, & Laudato, 2000; Kass & Raftery, 1995; Larkey, Smith, & Kadane, 1989). This confusion makes it difficult to draw meaningful conclusions about the rationality of people's judgments.

We investigate the ability of observers to track changes in dynamic environments. In contrast to research on the hot hand phenomenon, we use controlled dynamic environments where we know exactly how the statistical sequence was produced and at what time points the changes occurred. This knowledge allows us to improve on prior hot hand research because we know the true (but hidden) state of the world, and we can assess observers' rationality against this benchmark. The dynamic task environment also allows us to advance on prior Gambler's Fallacy research. In a typical Gambler's Fallacy experiment, observers are asked for predictions about a string of independent observations. The optimal response is to make predictions that are also sequentially independent, leaving just one option for observers to display sub-optimality: by inducing greater-than-zero sequential dependence in their predictions. With this constraint, any variability between observers forces sub-optimality to be observed in the average. In contrast, our dynamic environment includes changes in the hidden state at some trials but not others, which induces sequential dependence between observations. This can lead to sub-optimality in two distinct ways; observers can believe there are either more *or fewer* sequential dependencies than actually exist. When an observer detects more change points than really exist, they may behave sub-optimally because they react to perceived changes in the underlying environment that do not exist (e.g., a basketball coach who is prone to seeing hot hands where none are present). Conversely, when an observer detects too few change points, they may fail to adapt to short-lived changes in the environment. This tradeoff between detecting too few and too many change points has often been ignored in previous studies of change detection, which mostly assumed an *off-line* experiment where the task is to identify change points in a complete set of data that were observed earlier (see, e.g., Chinnis & Peterson, 1968, 1970; Massey & Wu, 2005; Robinson, 1964). However, real-world examples are invariably *online*: data arrive sequentially, and a detection response is required as soon as possible after a change point passes, before all the data have been observed. On-

Author's personal copy

line change detection is also important in clinical settings, for instance in identifying dorsolateral frontal lobe damage. The widely used Wisconsin Card Sorting Task (Berg, 1948) screens patients according to how often they make perseverative errors—that is, how often they fail to detect a change in the task environment, and continue to apply an outdated and sub-optimal strategy. Animal researchers have studied similar behavior in rats (e.g., Gallistel, Mark, King, & Latham, 2001). Rats take some time to detect and adjust to unsignaled changes in reinforcement schedules, but eventually return to probability matching behavior.

We develop a particle filter model as a psychologically plausible model for online change detection. Particle filters are Monte Carlo techniques for the estimation of the hidden state of a dynamically evolving system (see Doucet, de Freitas, & Gordon, 2001) and have recently been proposed as a general class of psychological models (Daw & Courville, 2007; Sanborn, Griffiths, & Navarro, 2006). They are attractive as models of human decision making because they require quite simple calculations, and do not require a long memory for past observations. Even with these limitations, particle filter models approach the statistically optimal treatment of the data when endowed with a sufficiently large number of particles—that is, the distribution of particles approaches the Bayesian posterior distribution across hidden states, conditional on all prior observations. In addition, by limiting the number of particles, the particle filter is able to mimic sub-optimal behavior observed in various tasks such as categorization (e.g., Sanborn et al.). Most interestingly, particle filters allow us to investigate what may be termed "conditionally optimal" models, in which statistically optimal algorithms are applied to incorrect initial assumptions. Such models help us to address the question of exactly where the sub-optimality in human information processing arises.

Using a simple change detection environment, we gather data from two experiments and compare participants' performance to predictions from a model based on particle filtering. The data reveal interesting trends, including the tradeoff between detecting too many and too few change points, and the model analyses shed light on how this tradeoff may operate. The model analyses also reveal that many participants behave in a "conditionally optimal" manner: data appear as if these participants apply rational statistical processes, but use incorrect estimates of certain environmental parameters. The close-to-optimal behavior of our participants stands in contrast to prior research on the fallibility of human decision making. We hypothesized that people have difficulties when they are asked for predictions about the future (as in gambling research, for example). Experiment 2 investigates the differences between predictions about the future vs. inferences about the past state of the environment. We observed a surprisingly large difference between these kinds of responses, and we use the particle filter model to provide a natural explanation for this difference.

1.1. Change detection environments

In our experiments, random numbers are presented to an observer, one number at a time. After each new value is presented, the observer is required to respond, either with an inference about the *mean* value of the process that is currently generating data, or with a *prediction* for the next value. Fig. 1 illustrates the particular set of distributions we used in our experiments, along with some example stimuli and two sets of example responses. Each stimulus was sampled from one of four normal distributions, all with the same standard deviation but with different means, shown by the four densities labeled "*A*" through to "*D*" in the upper right corner of Fig. 1. The 16 crosses below these distributions show 16 example stimuli, and the labels of the distributions from which they arose are shown just to the right of the stimuli; we call these the *hidden* states because these are unobservable to the participants in our experiments. The five uppermost stimuli were generated from distribution *A*. Most of these fall close to the mean of distribution *A*, but there are random fluctuations—for example, the third stimulus is close to the mean of distribution *B*. After the first five stimuli, the hidden state switches, and three new stimuli are produced from distribution *B*. The process continues with six stimuli then produced from distribution *D* and finally two from distribution *A* again.

There were two different kinds of questions that we asked participants. In the *prediction* condition, we asked about the *future*, requiring participants to predict where the next stimulus would fall. Participants made this response by clicking the mouse in one of four regions, defined by

Author's personal copy

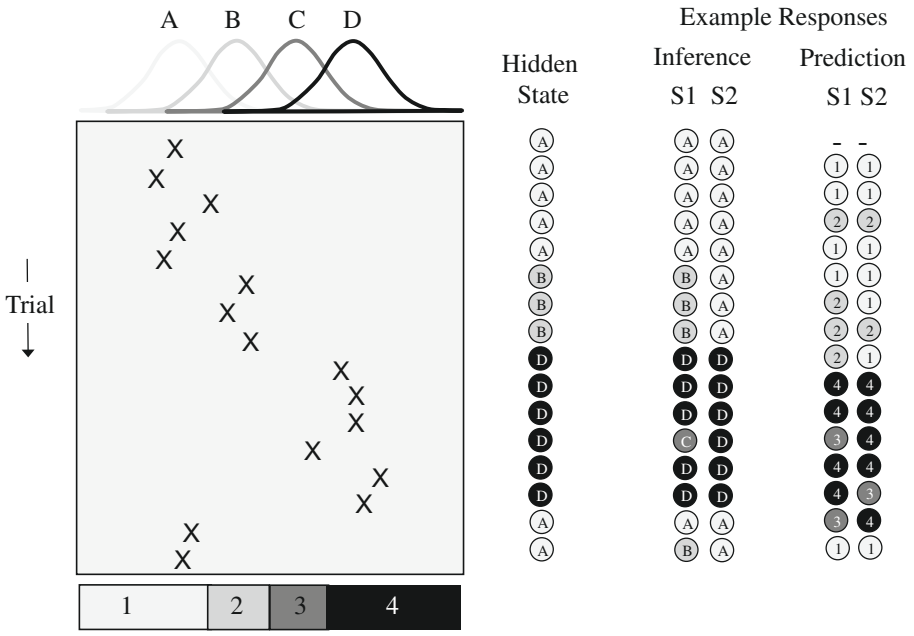

Fig. 1. An illustration of the change detection paradigm. Each stimulus is sampled from one of four normal distributions, but the selection of the hidden state (generating distribution) is not observed by the participant. At each timestep, there is a small probability that the hidden state is switched to a new distribution. Example data for two participants are shown on the right for two different tasks; an inference task where the goal is to identify the hidden state for the last stimulus, and a prediction task where the goal is to predict the most likely region where the next stimulus will appear. See text for additional details.

the crossover points of the four distributions and illustrated in Fig. 1 by the shaded rectangles in the lower left corner. Some example predictions for two participants (S1 and S2) are shown on the far right of Fig. 1. These data illustrate some general trends that will become important later—for example, the third stimulus is quite close to distribution *B*, and this causes both participants to predict that the next stimulus will fall in the corresponding region. The other kind of question we asked participants was about the *past*, which we call an *inference* response. In this condition, after each stimulus we asked the participants which distribution they think had generated it. In other words, with the inference task, we require participants to identify the hidden state of the generating process. Two example sets of responses are also shown for the inference task on the right side of Fig. 1.

We generated stimuli for the experiments using a hidden Markov model observed through a filtering distribution. On each trial, an observation was sampled randomly from one of four normal distributions depending on the hidden state. After every trial there was a probability α that the hidden state would change to one of the other three states, selected at random. The exact value of α was varied depending on the experimental condition. When α is small, most observations are sampled from the same distribution, requiring few change detections from the participant. With a large α, rapid fluctuations in the hidden states are possible, requiring participants to carefully interpret the changes in observed stimulus values. More formally, let the hidden state z_i represent which of the four generating distributions (labeled *A*, *B*, *C*, D in Fig. 1) was used to generate the stimulus used for trial *i*. We used the following algorithm to generate observable stimuli y_i for trials $i = 1, \ldots, T$ where T is the total number of trials:

(I) Set z_0 to a random sample from {*A*, *B*, *C*, *D*}.
(II) Repeat for trials $i = 1, \ldots, T$.
1. Set $z_i = z_{i-1}$.
2. Sample x uniformly from $[0, 1]$.
3. If $(x < \alpha)$ replace z_i with a random sample from $\{A, B, C, D\}/z_{i-1}$.
4. Sample stimulus y_i from a normal distribution with standard deviation σ and mean μ_{zi}.

Author's personal copy

For all experiments and model simulations, we chose mean values for the four distributions (μ_j) of {.2,.4,.6,.8} and set $\sigma = .1$ in order to have some overlap in the generating distributions and to insure that almost observations fall on the unit interval.

1.2. A particle filter model

Our task involved tracking the hidden state of an underlying process whose visible outputs are perturbed by noise. This is an important problem in many statistical and computational applications, commonly called the "filtering problem", in which particle filters have had great success. Particle filter models are a type of sequential Monte Carlo integration, and are the real-time (or online) analogs of standard Markov chain Monte Carlo integration methods (see, e.g., Doucet et al., 2001; Pitt & Shepard, 1999). For statistical analysis, the central advantages of particle filter models are their simple and efficient computational properties, and that they can be made to approximate the Bayesian solution for some problems, without the intractable integration problems that often arise in Markov chain Monte Carlo methods. Particle filters also make good candidates for psychological models because they can perform almost optimally, in the sense that they can arbitrarily closely approximate the posterior distribution over hidden states of the system, even though the required computations are simple enough to be plausible as a psychological process. The first application of particle filters to psychological data was by Sanborn et al. (2006), who showed that a particle filter model for categorization was able to capture both ideal observer as well as sub-optimal behavior simply by varying the number of particles. Particle filters also have similarities to existing, but less formal, models of cognition that have enjoyed considerable support, such as Kahneman and Tversky's (1982) "simulation heuristic". There are many methods for sampling from particle filter models (see, e.g., Doucet, Andrieu, & Godsill, 2000). We develop a model for the change detection task based on the method of direct simulation, which is the simplest of all particle filter methods and does not make unreasonable demands on the observer. The mathematical details of this algorithm are set out in Appendix A, but we provide an intuitive description in the following text, and an illustrative example in Fig. 2.

The observer begins with some number of *particles*, say P. Each particle represents a single hypothesis about the hidden state of a dynamically evolving system. In our task environment, a particle represents a single hypothesis for which distribution is currently generating observed stimuli. An initial set of particles for a $P = 8$ system is shown on the top row of the right hand side of Fig. 2. These initial particles have randomly distributed guesses—for example, three particles guess that the generating distribution is B, one particle that the distribution is D and so on. The first observation is generated from distribution C, and is illustrated by the uppermost cross on the shaded rectangle. After this observation, the particles are updated such that particles consistent with the observation are kept (or even multiplied) and those that are inconsistent are rejected. The first observation is most consistent with distribution C, but also somewhat consistent with B and D. This causes the two initial particles that hypothesized distribution A to be rejected—no arrows show these particles continuing to the next trial. After updating, only particles consistent with distributions B, C and D remain. The second observation (also drawn from distribution C) is most consistent with distributions C and D. This time, the update operation rejects the three particles from distribution B because they are inconsistent with the data. This process continues and the ensemble of P particles evolves over trials, and tracks the hidden state as it moves, illustrated by the histograms on the right side of Fig. 2. These show that the distribution of particles tracks the true hidden state (shown on the far left of the figure).

The particle filter model has two parameters that affect its performance, the number of particles (P) and an estimate of the frequency of changes in the generating process ($\hat{\alpha}$). The number of particles imposes a processing limitation, and affects the overall level of task performance: more particles lead to better performance, and in the limit of infinitely many particles the distribution of particles over states converges to the posterior distribution over states, conditional on *all* previous observations (it is in this sense that the model is optimal). It may at first seem psychologically implausible that an observer could maintain, say, 25 separate hypotheses about the state of the hidden process (i.e., a model with $P = 25$ particles). However, the information carried by those 25 particles is no more than the four

Author's personal copy

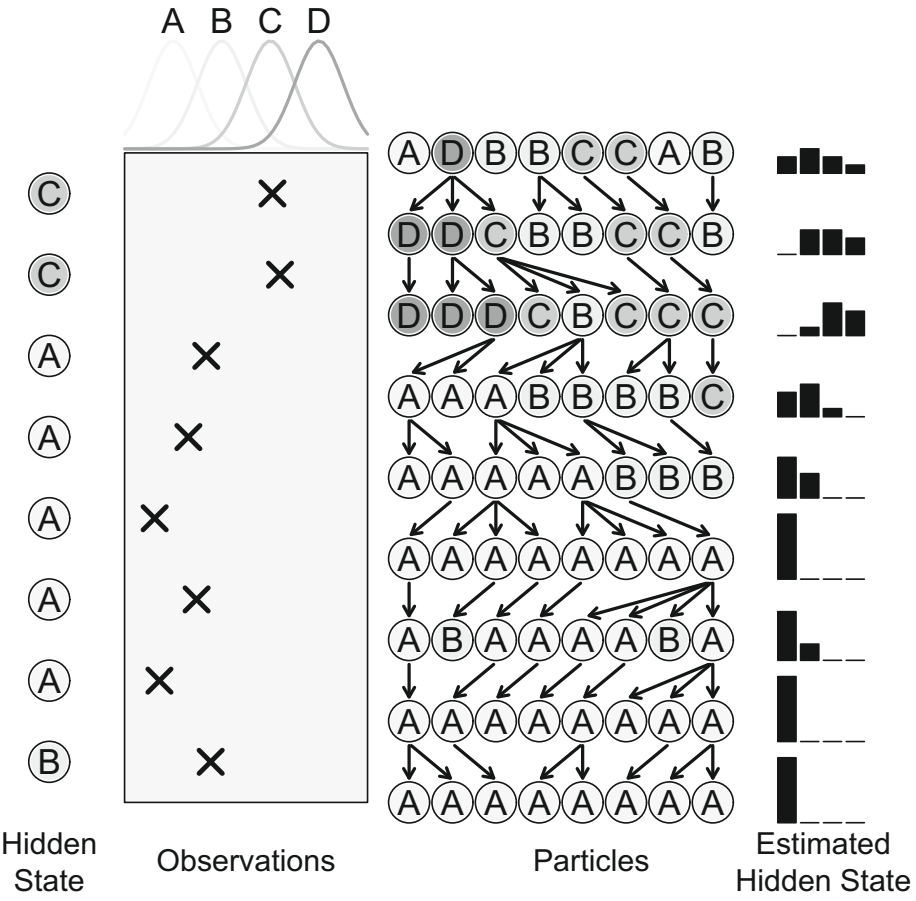

Fig. 2. Illustrative example of the particle filter algorithm.

counts of how many particles are associated with each of the four possible hidden states (*A*, *B*, C and *D*). Instead of thinking of the number of particles as a number of independent hypotheses, one might more plausibly consider it as the resolution with which the distribution over states is calculated.

The second quantity that affects the model's performance is an estimate for the value of α, which represents the observer's belief about how often the hidden state changes—we will call this estimate $\hat{\alpha}$. This parameter is used when particles are updated after each observation. When each particle is updated, it is transformed in a way that mimics changes in the hidden state: with probability $\hat{\alpha}$, the particle will change its state to one of the three other hidden states (selected at random), or will remain unchanged with probability $1 - \hat{\alpha}$. A statistically optimal model would use the true value of $\hat{\alpha} = \alpha$, but our participants cannot reasonably have known this. Even more fundamentally, our participants could not reasonably have known that we kept α fixed at a constant value on every trial. We incorporated into the model an assumption that leads to the Gambler's Fallacy, a belief that a change is more likely occur as more time passes without a change. In the actual task environment, this probability (α) is constant across time but we assume that observers are drawing inferences about a *different* environment where this is not the case. The model thus incorporates a Gambler's Fallacy, but only relative to task environments where the change probability is constant (or zero). In the model, each particle carries with it its own imperfect estimate $\hat{\alpha}$. This estimate begins at $\hat{\alpha} = 0$ when the particle is created (i.e., when it first changes to a new hidden state) and increases to $\hat{\alpha} = 1$ as the particle ages. Specifically, when a particle is k trials old we set $\hat{\alpha} = 1 - e^{-rk}$ where r is a rate parameter that measures how quickly the observer believes the hidden state of the system changes. Appendix A shows the model predictions as a function of both $\hat{\alpha}$ and the number of particles P. In our first experiment, we investigated the sensitivity of participants to changes in the true value of α. We examine how responses change when α is manipulated, and whether the observed response differences imply changes in the observers' underlying beliefs about the evolution of the hidden states (i.e., whether observers change their beliefs, $\hat{\alpha}$ in response to changes in α).

Author's personal copy

2. Experiment 1—Manipulating change frequency

In Experiment 1, we investigated the ability of participants to track the hidden state of a dynamically changing environment. After displaying each stimulus, we asked participants which of the four distributions was most likely to have generated that observation. We tested participants on this inference task only (the prediction task will be investigated in Experiment 2). We assessed performance in this task under three conditions, in which there was either a low, medium or high change probability (i.e., α value). We were interested to know whether participants were able to track changes in the hidden state and whether they were sensitive to the changes in the rate at which those changes occurred. Also, this experiment allowed us to investigate individual differences in the accuracy of change detection as well as the number of changes detected. We hypothesized there would be substantial differences in individual ability, with some individuals detecting too few or too many changes, leading to sub-optimal performance.

2.1. Methods

One hundred and three undergraduates from the University of California, Irvine, participated in Experiment 1. We generated stimuli using the algorithm described above, and illustrated these using a "tomato processing factory" in which cans of tomatoes were produced from one of four horizontally-separated "production pipes" at the top of the screen (to view a working example of the experiment, visit http://psiexp.ss.uci.edu/). The locations of these four pipes correspond to the mean values of four distributions that generate the stimuli. Using simple animations, tomato cans were produced one at a time from one of the four pipes. After each new stimulus appeared, all other stimuli on screen were moved downwards to make room, simulating a conveyor belt. No more than 15 stimuli were ever visible on screen at any time, with the lowest one falling off screen. There were four response buttons at the top of the screen, one above each pipe. Participants used these buttons to indicate their responses to the inference questions ("Which pipe generated the most recent stimulus?"). The experiment began with the participant reading through instructions that described the generating process and the random movement of the cans. After this, the first cans rolled out of the machine with all elements of the task visible. This "familiarization phase" lasted for 10 stimuli at the beginning of each block. The participant's task at this time was trivial, as they could directly observe which pipe had generated the most recent stimulus (the "hidden" state). The familiarization phase allowed us to identify and exclude participants who failed to pay attention. It also made the participants more familiar with the idea of inferring which pipe generated the stimuli, and illustrated that the stimuli did not always fall nearest to the pipe that had generated them. After 10 familiarization trials, a curtain covered the machinery that generated the stimuli, beginning the 40 trials of the decision phase. During this phase, the participant's task remained unchanged, but became more difficult since the hidden state was truly hidden: the only information available to the participant was the final location of the stimulus.

Participants completed one short practice block followed by 12 full blocks divided into four blocks in each of three conditions. The three conditions were defined by the frequency of changes in the hidden state: $\alpha = 8\%$, $\alpha = 16\%$ and $\alpha = 32\%$. The pipe used to generate each stimulus was either a repeat of the pipe used for the previous trial (with probability $1 - \alpha$) or a new pipe drawn randomly from the other three possibilities (with probability α). All four blocks of each condition occurred consecutively, but the order of the three conditions was counterbalanced across participants. We constrained the pseudo-random stimulus sequences to ensure that there was at least one change in the generating distribution during each familiarization phase. Importantly, we used the same stimulus sequence for corresponding blocks for all participants to reduce variability in comparisons across participants.

2.2. Results

The two central attributes of participants' responses were their accuracy (i.e., how often they correctly inferred which distribution had produced the data), and their variability; these measures are summarized in the left and right columns of Fig. 3, respectively. Participants averaged about 70% cor-

Author's personal copy

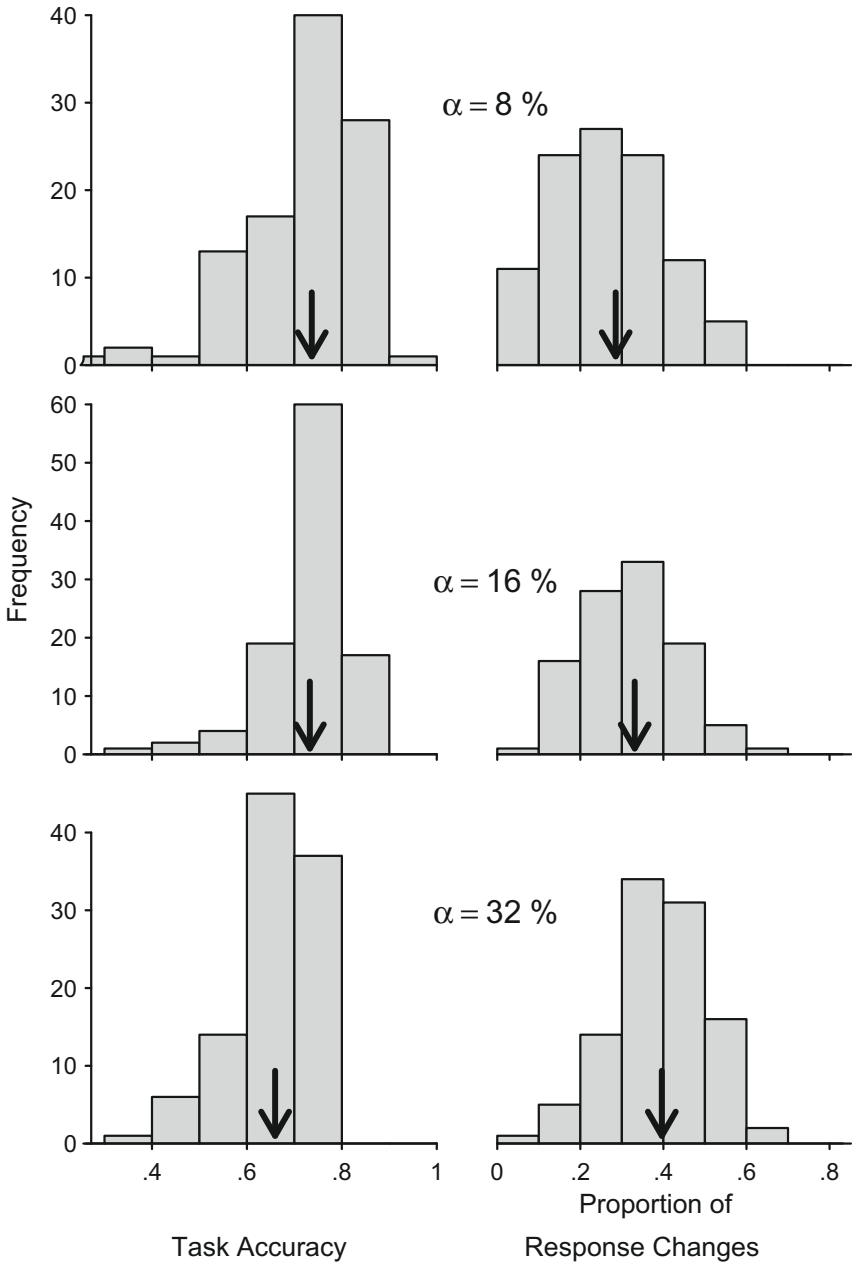

Fig. 3. Histograms of task accuracy and response variability across participants in Experiment 1. The arrows show the mean of each distribution.

rect responses, with a tendency for accuracy to decrease with increases in the proportion of changes in the hidden state (α). We summarized response variability by calculating the proportion of trials on which the participant provided a different response than they had provided for the previous trial, which provides a measure of how often the participant changed their belief about the hidden state. The histograms for response variability (right column of Fig. 3) show that variability increased as the proportion of changes in the hidden state increased. On average, participants made 11 response changes per block during the low frequency ($\alpha = 8\%$) condition, rising to 13 changes per block in the medium frequency ($\alpha = 16\%$) condition and 16 changes per block in the high frequency ($\alpha = 32\%$) condition.

Fig. 4 compares accuracy and response variability on a within-subjects basis, separately for the three conditions: low, medium and high α. Each participant is represented by a black dot in each of the three panels (the gray areas are predictions from the particle filter model, discussed next). The upper boundary of the accuracy data forms an inverted-U shape for all three conditions. This in-

Author's personal copy

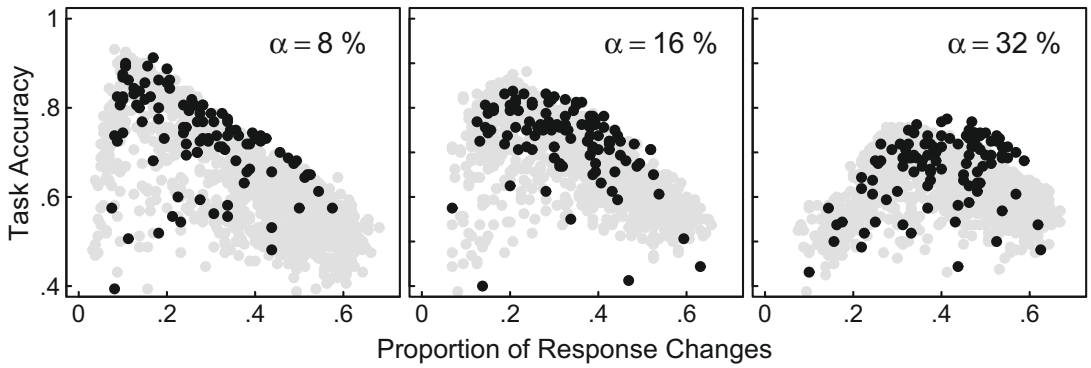

Fig. 4. Accuracy vs. response variability (the proportion of trials on which the participant changed their response). From left to right, the panels show data from the three conditions of Experiment 1: low, medium and high probabilities of a change (α) in the hidden state. Black dots show data, the gray dots shows predictions from the particle filter model, under all possible parameter values.

verted-U shape illustrates the tradeoff between subjects who were too quick to detect changes and those who were too cautions.

Our participants demonstrated increased response variability and decreased accuracy across the three conditions, as the frequency changes in the generating process (α) increased. This may suggest that participants were appropriately reactive to the experimental manipulation, but the situation may be more complicated. Increasing the frequency of changes in the generating process also increases the variability in the data, and so naturally leads to greater response variability. Put another way, even if participants were completely insensitive to the experimental manipulation of α, they would still exhibit increased response variability with increased α, just by virtue of tracking the (more variable) stimuli. The interesting question is whether participants demonstrated *enough* extra response variability to suggest they were sensitive to the experimental manipulation of α. This question can only be answered by model-based analyses.

2.2.1. Particle filter model analyses

The particle filter model maintains a set of particles that represent a distribution of beliefs about which of the four processes is currently being used to generate data—see, e.g., the histograms on the extreme right of Fig. 2. In Experiment 1, participants were forced to give just one response, not a distribution. This constraint is included in the model in the statistically optimal manner, by having the model return the *mode* of the particle distribution as its inference about the most likely generating distribution. This response is optimal in the sense that, if the particle distribution is an accurate approximation to the posterior distribution over hidden states, given the observed stimuli, then the mode will be the most likely state. Of course, given that we endowed the model with incorrect assumptions about the probability of change (recall that α was actually constant, but the model represented it as increasing with age) the model's predictions are only conditionally optimal. We use two different approaches to assess the model. We first take a global approach, in which we evaluate the model's predictions for *all* possible parameter settings, and compare the range of behaviors produced by the model with participants' behavior. Following that, we take a focused approach by estimating model parameters for each participant, and investigating the effects of the experimental manipulations on the parameter estimates.

The gray circles in Fig. 4 illustrate the entire range of behavior that can be predicted by the particle filter model when given the same stimuli as our participants.[1] We varied the number of particles from $P = 1$ to $P = 1000$ (predicted accuracy reached asymptote around $P = 250$) and we varied the rate at which α was assumed to change with particle age on a fine grid between $r = 0$ and $r = 1$. The model captures the observed variability in participant behavior in this task in two ways. Firstly, the model

[1] There is some random variability in the model predictions shown in Fig. 4, because of the finite sequence of data used—just four blocks of 50 trials for each condition, as seen by our participants.

Author's personal copy

successfully predicts data that were actually observed—almost all data fall inside the range of model predictions. Secondly, the model does not predict data that are wildly different from those that were observed—the gray circles do not generally fall very far from the data. The model also captures the important qualitative trends in Fig. 4, including the decreased accuracy and increased variability with increasing change probability (r). The upper boundary of the gray shading illustrates the conditionally optimal performance level that can be achieved for all different estimates of r (with large P). Even though the accuracy of the participants decreases as they detect too many and too few changes, the accuracy of many participants remains close to the top border of the gray shading. This indicates that many participants could be characterized as operating in a conditionally optimal manner: that is, optimal except that they held inaccurate ideas about α.

The model analyses provide insight into the question of whether our participants were sensitive to the experimental manipulation of α. We used the particle filter model to generate maximum-likelihood estimates of the parameters (P and $\hat{\alpha}$), separately for each participant. Estimating parameters using a particle filter model is a non-trivial problem for which various techniques exist, but none are completely satisfactory (e.g., Hürzeler & Künsch, 2001). We used a simple method based on our data summary graph (Fig. 4) but we also tried more sophisticated methods, with broadly similar results (see Appendix B for details). We constrained the parameter estimates to a single, fixed value of P for all three conditions for each participant, but allowed three independent values of r for the three experimental conditions, reflecting the experimental manipulation of change rate. The parameter estimates revealed that participants were indeed sensitive to manipulations of the underlying frequency of changes in the data generating process. Fig. 5 shows histograms of the r estimates across participants, separately for the three experimental conditions. The r estimates were smallest in the $\alpha = 8\%$ condition (mean $r = .18$), increasing for the $\alpha = 16\%$ condition (mean $r = .21$) and the $\alpha = 32\%$ condition (mean $r = .28$). The estimated number of particles for each participant ranged from $P = 1$ to $P = 400$ with a mean of $P = 56$.

2.3. Discussion

The remarkable finding from Experiment 1 was that many participants performed in a conditionally optimal manner—that is, they behaved as if they applied statistically optimal calculations to the data, although based on an incorrect assumption about the rate of changes in the hidden state of the system. The reader may wonder why many of our participants performed so well, when data from other paradigms reliably demonstrate the fallibility of human decision making. Our dynamic data generating process may be similar to "real world" environments than static (i.i.d.) processes, resulting in better performance. A second feature that separates our paradigm from others is the question posed to participants. We asked our observers to identify which of the four stimulus distributions generated the most recent stimulus. This question asks about the *past*, requiring participants to make inferences

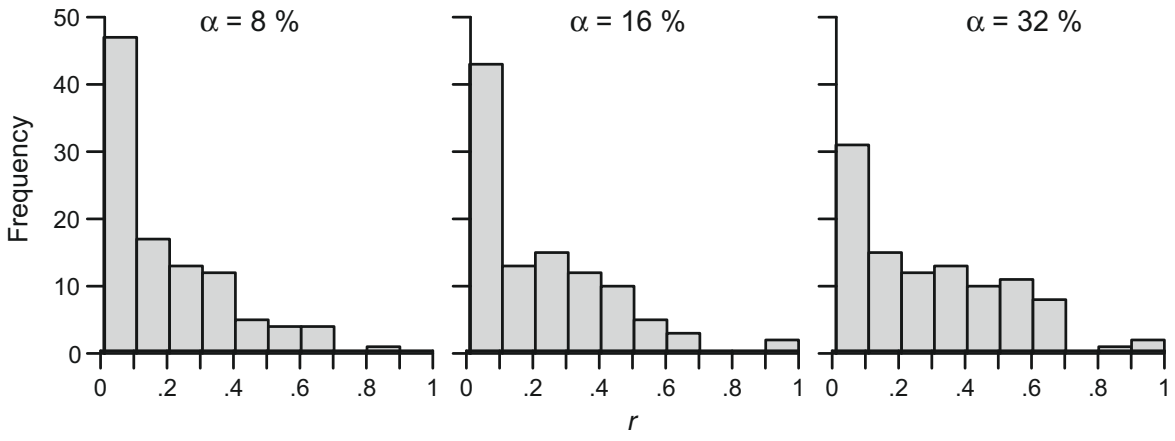

Fig. 5. Estimates of the r parameter from the particle filter model. The estimates increase as the true value (α) increases from 8% to 32%.

Author's personal copy

about prior states of the world using existing data. Most other analyses of human decision making ask instead about the *future*, requiring participants to make predictions about the outcomes of gambles, for example (e.g., see Kahneman & Tversky, 1973). It is possible that people treat questions about the past and about the future very differently, even when there is no statistical reason to do so. Similarly, Jones and Pashler (2007) have shown that making predictions about the future is not an especially privileged task. Indeed, they found that when similar questions were asked about the future (predictions) and the past ("retrodictions"), participants never performed better when making predictions than retrodictions. As an additional benefit, Experiment 2 also allows us to gather support for the way in which we included the Gambler's Fallacy belief in our particle filter model—this mechanism allows the model to accommodate a complicated pattern of results regarding response runs and alternations in Experiment 2, with no parameter adjustments.

3. Experiment 2—Past and future

Performance in Experiment 1 was very good when compared with classic findings from the literature, which detail the many flaws in human decision making. Experiment 2 was designed to test a hypothesis that may explain this difference. In classic decision making paradigms, participants are almost always asked to make predictions about the future—for example, to place bets on gambles. In Experiment 1, we asked participants about the past instead, by asking them to infer what underlying distribution had generated an observation that they had already seen. Experiment 2 uses the same paradigm, and even the same stimuli, to compare performance on predictions about the future with inferences about the past.

3.1. Methods

Another (different) 63 undergraduates participated in Experiment 2, which was very similar to Experiment 1. In Experiment 2 we used the same frequency for changes in the hidden state throughout the experiment, an $\alpha = 10\%$ chance of change on each trial. Once again we used 12 blocks of 50 trials, but this time they were divided into two conditions with six blocks each—all six blocks for each condition always occurred consecutively, with the order of the two conditions counterbalanced between subjects. The "inference" condition was identical to Experiment 1—the observer was asked about the *past* ("Which distribution generated the most recent stimulus?"). In the six blocks of the "prediction" condition, we required responses about the *future* instead ("Where will the next stimulus fall?"). Participants made their prediction responses by clicking the mouse in one of four regions of the output space. These response regions corresponded to the most likely stimulus locations from the four distributions, as shown by the shaded rectangles in the lower left of Fig. 1.

We used exactly the same pseudo-random sequence of hidden states and stimulus locations for the inference and prediction conditions. Given the same stimuli, the inference and prediction questions are mathematically identical,[2] and there was no statistical reason to give different responses in the two conditions (the best possible prediction for the next stimulus is just the region corresponding to the distribution that generated the last stimulus). This property allowed us to compare past and future responses directly, attributing any changes in behavior to task instruction only. The sequences were long enough (300 random normal deviates) that it seemed unlikely participants would be able to remember the sequence from the previous block.

3.2. Experiment 2—Results

We first illustrate some raw data, and then move on to data summaries and model-based analyses. Fig. 6 shows data from three representative participants in the fifth block of both the inference task

[2] As long as $\alpha < .75$, the distribution most likely to generate the next observation is whichever distribution generated the previous observation. The likelihood function for the next observation is a mixture of the four distributions, with mixing probabilities given by $\hat{\alpha}$ and $1 - \hat{\alpha}$, but the mode of this distribution will be unchanged as long as $\hat{\alpha} < .75$.

Author's personal copy

and the prediction task. The top rows use gray shading to illustrate the generating distributions (1...4 from black to white) for the 40 decision-phase trials. The second-to-top rows use the same shading to show where the stimuli actually fell: note that the stimuli were identical for the inference and prediction blocks. The three lowest rows show responses from three representative participants (S1, S2 and S3) using the same shading system.

Fig. 6 illustrates data from three participants who performed the task quite well. At trial number 32, the hidden state at the top of the figure changes from #1 (black) to #2 (dark gray). The three participants all changed their responses appropriately, around two trials later. Participants were also responsive to outliers. For example, hidden state #1 was used from trial 10 until trial 32, but several outlying stimuli were randomly produced. Two of these outliers occurred together on trials 24 and 25, and two of the three representative participants briefly changed their responses about the generating distribution around trial number 26.

For the three representative participants in Fig. 6, responses were less variable in the inference task than the prediction task—there are more response changes on the right side of Fig. 6 than the left. This was surprising, given that there was no statistical reason to give different responses under the two conditions. We further investigated the extra response variability in the prediction task by plotting data summaries for all participants, similarly to Experiment 1. Fig. 7 graphs accuracy vs. response variability, in the same format as used for Fig. 4. Data from the inference task were very similar to those from Experiment 1, which was to be expected given that the procedure was almost identical. Accuracy was lower in the prediction task. This was also unsurprising, as the prediction task was more difficult than the inference task: an accurate response required not only knowing which distribution will generate the next observation, but also being lucky enough to have that observation fall within the response region associated with its distribution. However, there was no reason to expect differences in response variability. The stimulus sequence used in the two tasks was identical, so the best strategy in the prediction task was to replicate responses from the inference task. Nevertheless, Fig. 7 shows that responses were more variable in the prediction task than the inference task.

The within-subject differences between inference and prediction tasks are shown more clearly in Fig. 8. The left panel graphs each participant's accuracy in the prediction task against their accuracy in the inference task. All data lie below the diagonal ($y = x$), indicating that accuracy was *always* lower in the prediction than inference condition. The center panel shows a similar analysis for response variability. Most participants (48 out of 63) made more variable responses during the prediction task than the inference task: on average, participants made 15 response changes per block during the future task compared with 11 per block during the past task. The right hand panel most directly illustrates the needless sub-optimality of participants in the prediction task. This graph shows the accuracy of responses in the prediction task, compared with how accurate those same participants would have been, had they simply re-produced the exact sequence of responses they gave for the inference task. Out of

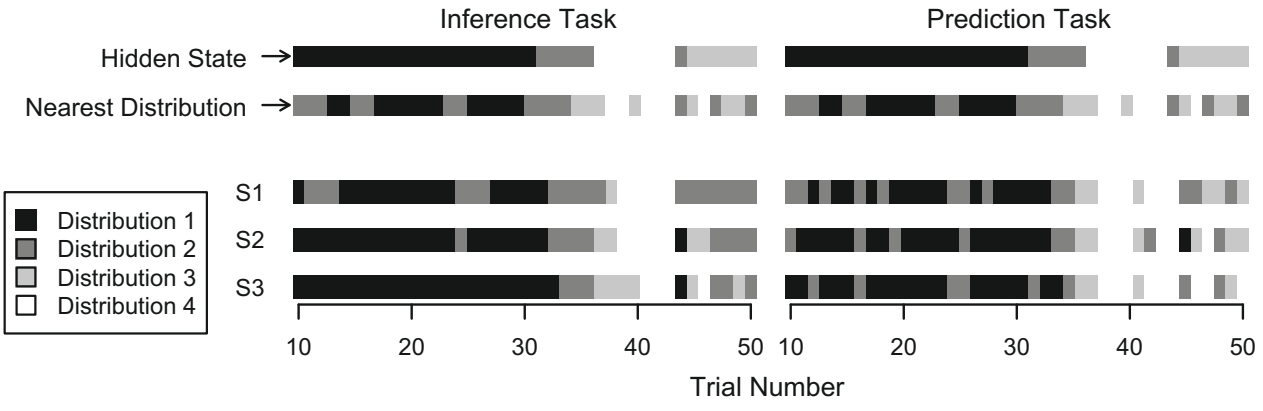

Fig. 6. Stimuli and data from one block of the inference task (left side) and one block the prediction task (right side). The top rows show the hidden state on each trial, represented by different shading—see legend. The next row shows which production distribution was closest to the randomly-displaced observation. The three lower rows show responses from three participants.

Author's personal copy

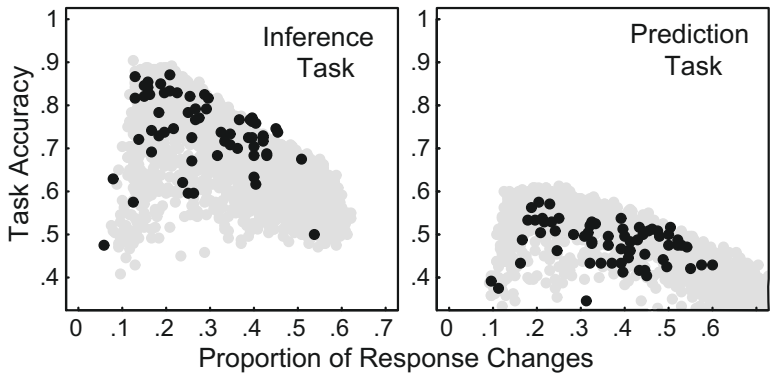

Fig. 7. Accuracy vs. the proportion of trials on which the response changed. The data are from the inference condition (left) and the prediction condition (right) from Experiment 2.

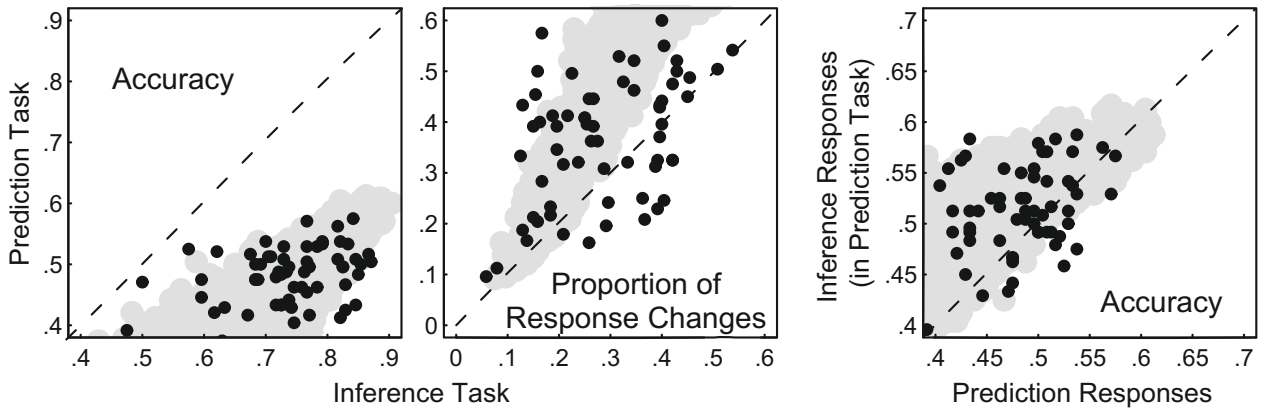

Fig. 8. The left panel compares task accuracy in the inference condition vs. the prediction condition, for each participant, and the center panel makes a similar comparison for response variability. The right panel shows how accurate each participant was in the prediction task (*x*-axis) and how accurate they *would* have been, had they simply replicated their own response sequence from the inference condition (*y*-axis).

the 63 participants, 46 of them (73%) would have been more accurate in the prediction phase if they had re-produced their responses from the inference phase.

3.2.1. Model analyses

There were striking differences in both accuracy and response variability between our prediction and inference conditions. The particle filter model, as outlined so far, only describes performance in the inference condition, characterizing this performance as statistically optimal, conditional on an incorrect assumption about the frequency of changes in the hidden state. In our analysis of data from Experiment 2, we impute powerful support for this model because it successfully accounts for the surprising differences between inference and prediction conditions with no modifications whatsoever—the data patterns arise as natural consequences of the model architecture.

Recall that the particle filter model produces an estimate of the posterior probabilities of the hidden states, represented as a distribution of particles across the states. To maintain (conditional) statistical optimality, responses to inference questions were given by the mode of the particle distribution, in both Experiment 1 and the inference condition of Experiment 2. We continued this theme to model responses in the prediction condition, generating responses from the particle distribution in the statistically optimal manner. Prediction questions effectively ask the observer about the hidden state one trial ahead in the future. The optimal response in this case is to evolve the particles one trial further into the future and respond with the mode of that evolved distribution. Evolving particles one step into the future means simply running the usual algorithm that updates the particles

Author's personal copy

after each observation, but without having the observation itself. In detail, step #3 of the algorithm as specified in Appendix A is replaced by one that accepts *all* proposed particles, rather than just those that are consistent with the new observation. The particles from the evolved distribution are thus identical copies of the existing particles, except that each particle has a chance of changing to one of the other generating distributions, with probability $1 - e^{-rk}$, where r is the estimate of change rate and k is the age of the particle (in trials). These new particles represent the best prediction for the state of the system one trial into the future, given no observed data. This procedure adds response variability, especially when there are low numbers of particles, because the new particles sometimes differ from the current ones, and this response variability causes decreased task accuracy, as observed.

The gray shading in Fig. 8 shows predictions from the particle filter model, for a fine grid of values on $1 \leqslant P \leqslant 1000$ and $0 < r < 1$. The model fits the data well, even though it has just two parameters, fixed across task conditions. The mean estimated parameters for this model were $P = 70$ and $r = .12$. The model successfully captures all of the behavioral patterns illustrated in Figs. 8 and 9, most importantly the decreased accuracy and increased variability in the prediction than inference condition.

3.2.2. Revisiting the Gambler's Fallacy

When asked for predictions about the next stimulus, participants produced greater response variability than when they were asked for inferences about the last stimulus. Increased response variability for predictions is reminiscent of the Gambler's Fallacy—a tendency to predict changes after a run of similar stimuli. We investigated this link by examining the effect of "streaks" of repeated stimuli on responses (for a similar approach, see Burns & Corpus, 2004). Fig. 9 graphs the probability that the previous response will be repeated on the current trial against the length of the streak—that is, against is the length of the current streak of outcomes. Successive outcomes were classified as the same (and the streak classified as continuing) whenever they fell within the same one of the four output regions shown in Fig. 1.

The analysis in Fig. 9 illustrates a qualitative difference between data from the inference and prediction conditions. For short runs of similar stimuli (streak lengths of 1–4) the probability of repeating the prior response increases as the streak length increases. This makes statistical sense, and is true in both the inference and prediction conditions. As the streak gets longer than four trials, the probability of repeating the prior response continues to increase in the inference condition. However, in the prediction condition, this probability begins to *decrease* with increasing streak length. This pattern is

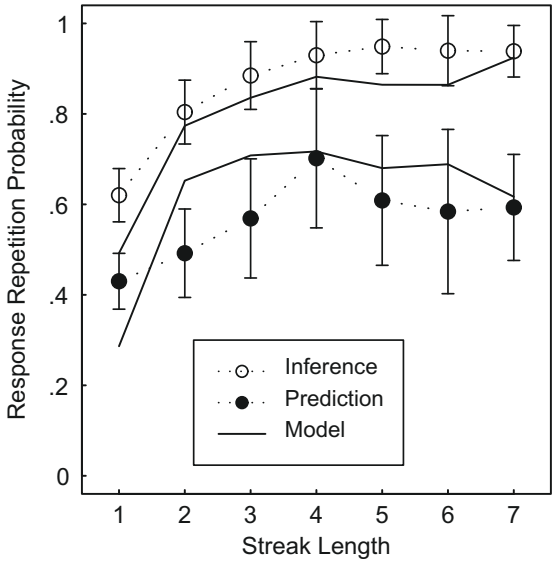

Fig. 9. The probability of repeating the previous response (y-axis) against the streak length: the number of trials for which the same outcome has been observed. Circles show averaged data, with 95% confidence intervals assuming binomially distributed probabilities across participants.

Author's personal copy

indicative of a Gambler's Fallacy: after a sufficiently long streak of similar outcomes, participants predict an increasing probability of a change.

The solid lines in Fig. 9 show predictions from the particle filter model. These predictions were generated in the same way as the data—separately for each participant (using best-fitting parameters) then averaged across participants. The model underpredicts the probability of repeating the prior outcome for very short streaks, but otherwise captures the data well. Critically, the model captures the qualitative difference between inference and prediction conditions, ascribing this difference to the lack of observed data in the prediction condition: predictions are made about a future stimulus, that has not yet been observed, while inferences use an existing, observed stimulus. In general, the model predicts decreased response repetition with longer streaks because particles get older as the streak length extends. As particles age, they have increased estimates of the underlying change probability of the system (according to $1 - e^{-rk}$), so when new particles are generated they are more likely to differ from the old ones as streak length increases. This means that the mode of the new particles becomes more likely to differ from the mode of the old particles as their ages increase, leading to the observed drop in response repetition with increased run length. The same pattern does not occur in the inference condition because there new particles are constrained by the observed data—new particles that differ from the old particles are unlikely to survive because they provide a poor match to the observed stimulus.

4. General discussion

4.1. Links with prior work

Modern Monte Carlo techniques provide a natural way to model inference in probabilistic environments, including many decision making tasks commonly used in psychological research. Standard Markov chain Monte Carlo (MCMC) methods, and the newer sequential Monte Carlo (SMC) methods, are both useful solutions to some of the associated computational problems, and variants of each have been proposed as psychological models (e.g., Brown & Steyvers, 2005; Sanborn et al., 2006). While both MCMC and SMC are equally appropriate analyses from a statistical viewpoint, it seems that SMC methods such as particle filters have several advantages as models of human cognition. Firstly, SMC algorithms place lighter and more plausible computational demands on the observer. SMC methods do not require the observer to keep a memory for a long chain of previous events. A related advantage for SMC methods is their ease of application to online experiments—when observations arrive sequentially, and a response is required after each observation. SMC methods are naturally designed for such sequential paradigms, and employ quite simple incremental update algorithms between observations. In contrast, many of the standard MCMC approaches require extensive re-calculation between each observation, making them computationally inefficient and untenable as psychological process mechanisms.

A final advantage of SMC approaches to cognitive modeling is the ability to model both optimal and sub-optimal behavior within the same framework. Standard Bayesian approaches to statistical inference based on MCMC naturally provide optimal performance, under the assumption of accurate knowledge of the problem structure. When endowed with a large number of particles, the particle filter models provide the same optimality as the MCMC techniques, but they can also accommodate sub-optimal behavior, when the number of particles is small. Both approaches allow investigation of interesting questions of conditional optimality. That is, when behavior is observed to be sub-optimal, it is interesting to ask what causes the decreased performance. Models such as ours make very specific predictions. For example, our model makes the strong assertion that all the observer's computations are statistically optimal, with one exception—an incorrect assumption about the rate of changes in hidden state of the system.

Animal research has shown that rats and pigeons, at least, are able to rapidly detect changes in reinforcement schedules and then to adjust their behavior in an almost-optimal manner (e.g., Gallistel et al., 2001). Such optimality is surprising and raises questions about the mechanisms that underlie performance. Daw and Courville (2007) proposed that particle filters are a plausible mechanism for

Author's personal copy

change detection in animals, and further suggest that particle filter models are able to capture individual participant behavior. Estes (1984) and Massey and Wu (2005) examined related paradigms with human observers, as in our experiments. In both cases, participants were reactive to changes in environmental parameters, but performed sub-optimally due to delays in detecting changes. Our research extends this work in two important ways. Firstly, Massey and Wu used only *one* change in the observations for each block, whereas we allowed many. This difference is important, because participants do not have to manage the complex tradeoff between detecting too many and too few changes—the number of changes is already known. Secondly, we extend prior work through our particle filter analyses, which allow us to address previously intractable questions. For example, in Experiment 1 we were able to show that participants were reactive to second-order changes in the task environment—changes in the frequency of changes.

5. Conclusions

As in our prior work (Brown & Steyvers, 2005; Brown, Steyvers, & Hemmer, 2007), we have shown that people manage a complex tradeoff between over- and under-reactivity when detecting changes in random sequences, with implications for real-world decision making and process monitoring applications. Experiment 1 demonstrated that many participants were able to perform a change detection and inference task in a conditionally optimal manner. In Experiment 2 we found that response accuracy decreased dramatically when participants were asked to make predictions about the *future* rather than giving inferences about the *past*. The decreased accuracy was characterized by the introduction of sequential dependencies in response sequences, with run length effects similar to those observed in traditional Gambler's Fallacy research. Nevertheless, the data were well described by the same conditionally optimal model as before. These results suggest that well-known and robust phenomena such as the Gambler's Fallacy may be absent, or at least much reduced, if questions are phrased to avoid the making of predictions.

Previously (Steyvers & Brown, 2005), we have modeled change detection processes using statistically optimal but psychologically implausible Bayesian inference algorithms. Here, we have shown that a particle filter model can give a good account of all our data. The particle filter model can be smoothly adjusted between very poor performance to statistically optimal performance. Many of our participants operated as if they were conditionally optimal observers (with a large number of particles), using optimal operations applied to an incorrect idea about the frequency of changes in the hidden state. The particle filter analyses also suggested a natural explanation for the difference in performance between prediction and inference task.

Acknowledgments

This research was supported in part by Australian Research Council Discovery Grant DP0878858, and a Queen Elizabeth II Research Fellowship for Brown, and by award FA9550-07-1-0082 from the Air Force Office of Scientific Research.

Appendix A

A.1. Particle filter based on direct simulation

Let U and φ be the uniform and normal density functions, respectively. Suppose that there are K hidden states, P particles, and that the observation on trial i, say y_i is normal with standard deviation σ and mean μ_j, where j is the hidden state. The algorithm is initialized on trial $i = 0$ with particles $u_0 = \{u_{0,1}, u_{0,2}, \ldots, u_{0,P}\}$, where each element is one of the integers $1, \ldots, K$. Each particle also carries with it an "age" initially set to zero: $a_0 = \{a_{0,1}, a_{0,2}, \ldots, a_{0,P}\}$, with $a_{0,i} = 0$. Let r be a parameter governing the expected rate of changes in the hidden state. Then on each trial i set a counter $p = 0$ and then repeat the following:

Author's personal copy

1. Sample z randomly from $\{u_{i,1}, u_{i,2}, ,\ldots, u_{i,P}\}$.
2. Sample $x_1 \sim U[0,1]$. If $x_1 < (1 - e^{-rk})$, where k is the age of the sample particle $(a_{i,p})$ then generate a new particle by sampling $z \sim \{1,2,,\ldots,K\}/\{z\}$ and setting $a_{i,p} = 0$. Otherwise, keep the old particle $(z = u_{i,p})$ and increment its age $a_{i,p} = a_{i,p} + 1$.
3. Sample $x_2 \sim U[0,1]$. If $[\varphi(y_i|\mu_z,\sigma) > x_2]$ then $p = p + 1$ and $u_{i,p} = z$.
4. If $p = P$ stop, else return to 1.

This method of direct simulation mimics the data generation process, but creates an evolving set of particles that approximates the desired posterior distribution.

Fig. A1 shows how these two parameters affect task accuracy (y-axis) and response variability (x-axis, the proportion of trials on which the predicted response changes from the previous trial). Each of the three lines shows model predictions for a different number of particles, $P = 2$, 6 and 150. As expected, increasing the number of particles leads to improved accuracy. Each line sweeps out a range of values corresponding to different values of r in the range zero to one. A very small value of r represents the belief that the hidden state of the system does not change very often, so the $\hat{\alpha}$ estimate associated with each particle increases very slowly from zero towards one as the particle ages. This leads to very few predicted changes in the model's responses, giving points on the left hand side of the graph. These points also have poor predicted accuracy, regardless of how many particles are used. A small value of r causes decreased performance because it governs how new particles are generated from old ones. Just as in the data, particles remain the same from trial to trial, except for a probability $\hat{\alpha}$ of changing to one of the other three locations. Similarly, a large estimate of r instantiates the belief that the hidden state changes very rapidly, which causes the model to predict very large response variability, which also causes poor accuracy.

Appendix B

B.1. Parameter estimation details

We explored two ways to calculate parameter estimates, both of which produced quite similar results. In the main text we report results only from the simpler method, which is both easier to communicate and more robust. The simpler method involved the comparison of critical summaries of the

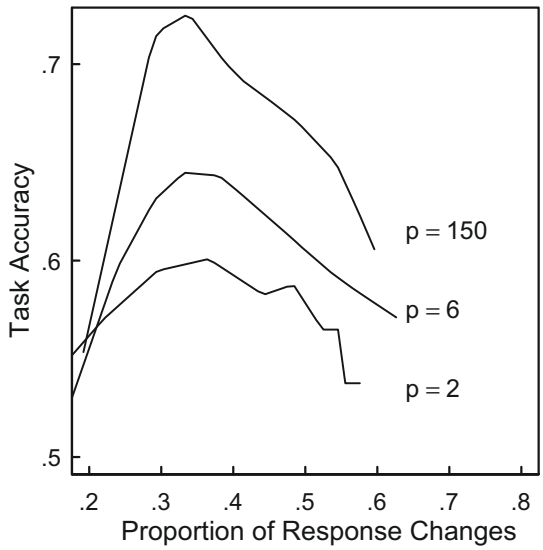

Fig. A1. Predicted task accuracy vs. predicted variability in responses (i.e., the proportion of trials on which the predicted response from the model changes). The three curves represent model predictions for three different numbers of particles ($P = 2$, 6, 150) showing that more particles leads to greater accuracy. Each curve sweeps out the model's predictions for all estimates of $0 < r < 1$. Small estimates of r instantiate a belief in very few changes in the hidden state of the system, so few response changes are predicted, and vice versa.

Author's personal copy

data with similar summaries calculated from model predictions. In Experiment 1, each subject's data was characterized by overall accuracy and by the proportion of response changes (separately for three experimental conditions). We calculated these quantities also for predictions from the particle filter model, for many different parameter settings. Then, the best-fitting parameters for a subject were calculated by finding those parameters that produced the closest match between the predicted and observed quantities. This has the effect of choosing those parameters that produced the closest match between the observed and predicted positions in Fig. 4 (in the sense of Euclidean distances). We used an analogous approach for Experiment 2, but we included two lots of data summaries for each participant, corresponding to Figs. 7 and 8.

We also calculated parameter estimates using a more complicated method, similar to maximum-likelihood estimation. Maximum-likelihood estimation using particle filters can be computationally difficult (see, e.g., Cappé, Doucet, Lavielle, & Moulines, 1999). In particular, the probability of an observed sequence of responses under a predicted sequence of responses is almost always zero, as there is almost always at least one trial with a mismatch between observation and prediction. For all combinations of P and r parameter values in a fine grid on $1 \leqslant P \leqslant 1000$ and $0 < r < 1$, we generated a complete set of responses for each of 200 pseudo-subjects. We then calculated the mean likelihood of the human responses over all pseudo-subjects separately for each parameter vector, and chose those parameter estimates that maximized the likelihood of the observed data separately for each participant. We solved the log-zero problem by setting the probability of each response equal to the frequency of particles for that response, plus a smoothing amount (5%), and normalizing. We then used a Luce choice rule (with unit temperature) to transform these frequencies to response probabilities.

References

Adams, R. M. (1992). The "hot hand" revisited: Successful basketball shooting as a function of intershot interval. *Perceptual and Motor Skills, 74*, 934.

Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association, 88*, 1175–1183.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15–22.

Brown, S. D., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 587–599.

Brown, S. D., Steyvers, M., & Hemmer, P. (2007). Modeling experimentally induced strategy shifts. *Psychological Science, 20*(1), 40–46.

Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: Gambler's Fallacy" versus "hot hand". *Psychonomic Bulletin & Review, 11*, 179–184.

Cappé, O., Doucet, A., Lavielle, M., & Moulines, E. (1999). Simulation-based methods for blind maximum-likelihood filter identification. *Signal Processing, 73*(1–2), 3–25.

Chatterjee, S., Yilmaz, M. R., Habibullah, M., & Laudato, M. (2000). An approximate entropy test for randomness. *Communications in Statistics: Theory and Methods, 29*(3), 655–675.

Chinnis, J. O., & Peterson, C. R. (1968). Inference about a nonstationary process. *Journal of Experimental Psychology, 77*, 620–625.

Chinnis, J. O., & Peterson, C. R. (1970). Nonstationary processes and conservative interference. *Journal of Experimental Psychology, 84*, 248–251.

Daw, N. D., & Courville, A. C. (2007). The rat as particle filter. *Advances in Neural Information Processing, 20*, 369–376.

Doucet, A., de Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. New York: Springer.

Doucet, A., Andrieu, C., & Godsill, S. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing, 10*(3), 197–208.

Estes, W. K. (1984). Global and local control of choice behavior by cyclically varying outcome probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(2), 258–270.

Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review, 13*, 737–741.

Gallistel, C. R., Mark, T., King, A., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 354–372.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108*, 33–56.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

Hürzeler, M., & Künsch, H. R. (2001). Approximating and maximizing the likelihood for a general state-space model. In A. Doucet & N. De Freitas (Eds.), *Sequential Monte Carlo methods in practice*. New York: Springer.

Jones, J., & Pashler, H. (2007). Is the mind inherently forward-looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review, 14*, 295–300.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 263–291.

Author's personal copy

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.

Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the hot hand. *Chance, 2*(4), 22–30.

Massey, C., & Wu, G. (2005). Detecting regime shifts: The causes of under- and overreaction. *Management Science, 51*(6), 932–947.

Pitt, M. K., & Shepard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association, 94*(446), 590–599.

Robinson, G. H. (1964). Continuous estimation of a time-varying probability. *Ergonomics, 7*, 7–21.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society.*

Sundali, J. J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's Fallacy. *Judgment and Decision Making, 1*, 1–12.

Steyvers, M., & Brown, S. D. (2005). Prediction and change detection. *Advances in Neural Information Processing Systems, 18*, 1281–1288.

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review, 12*, 409–441.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331–350.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology: General, 134*, 117–123.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review, 11*, 579–615.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General, 134*, 108–116.