Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Systems Research 6 (2005) 5–16
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Streak biases in decision making: data and a memory model

Action editor: Christian Schunn

Erik M. Altmann *, Bruce D. Burns *

Department of Psychology, Michigan State University, East Lansing, MI 48824, USA

Received 16 September 2004; accepted 16 September 2004

Abstract

Streaks of past outcomes, for example of gains or losses in the stock market, are one source of information for a decision maker trying to predict the next outcome in the series. We examine how prediction biases based on streaks change as a function of length of the current streak. Participants experienced a sequence of 150 flips of a simulated coin. On the first of a streak of heads, participants showed positive recency, meaning that they predicted heads for the next outcome with a greater-than-baseline probability. As streak length increased, positive recency first decreased but then increased again, producing a quadratic trend. We explain these results in terms of outcome-prediction processes that are sensitive to the historical frequency of streak lengths and that make heuristic assumptions about changes in bias of the outcome-generating process (here, the coin). An ACT-R simulation captures the quadratic trend in positive recency, as well as the baseline heads bias, in two experimental conditions with different coin biases. We discuss our memory-based model in relation to a model from the domain of economics that posits explicit representation of an "urn" from which events are sampled without replacement.

Keywords: Decision making; Memory; Cognitive simulation; ACT-R; Hot hand; Gambler's fallacy; Streaks

1. Introduction

A streak of repeated outcomes can be an important source of information for decision makers trying to predict the outcome of the next event. For example, asked to predict whether the global average temperature will increase or decrease next year, a decision maker who has access to historical records may predict an increase simply on the basis of past trends. A bias of this form, in which a streak of past outcomes is taken as evidence that the next outcome (or measurement) will be in kind, is often referred to as *positive recency*. In terms of

* Corresponding authors. Tel.: +1 517 353 4406; fax: +1 517 353 1652.
E-mail addresses: ema@msu.edu (E.M. Altmann), burnsbr@msu.edu (B.D. Burns).

1389-0417/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2004.09.002

the decision-maker's causal reasoning, one could imagine that a long streak of temperature increases induces a belief that an underlying causal mechanism is at work (warming due to greenhouse gases, for example). This causal mechanism will then govern changes in next year's measurement as well (modulo extraneous variance).

Of course, whether the decision maker makes this or the opposite prediction will depend on precisely how he or she represents the mechanism producing the streak. If the decision maker happens to work for the current Bush administration, he or she may well understand historical increases in the global temperature in terms more similar to the well-known gambler's fallacy, in which a gambler takes a streak of undesirable outcomes as evidence that his or her luck will soon change. More generally, a bias of this form, in which a streak of outcomes is taken as evidence that the next outcome will be opposite, is often referred to as *negative recency*. In the case of gambling, each gamble is an independent event, so there is no causal mechanism linking the outcomes (hence the fallacy). However, there are situations in which negative recency is a rational bias, namely when outcomes are sampled without replacement. For example, rats and other foraging animals have a bias against returning to the location where they found food on a previous trial (Olton, 1978). The underlying causal model is, presumably, that food at a given location is depleted before it is replaced; if this model is correct, then negative recency is adaptive.

In this paper we examine how recency biases change as a function of streak length. That is, we are interested in how the length of a streak, up to and including the most recent outcome, affects the decision maker's prediction concerning the next outcome. An experimental paradigm appropriate for addressing such issues involves a two-choice prediction task (similar studies complete with simulations include, e.g., Lebiere, Gray, Salvucci, & West, 2003; Lovett, 1998). In the standard experiment with this kind of paradigm, the participant is asked on each trial to predict the outcome of an event such as a coin flip. The "coin" typically has a bias toward one outcome or the other, of which the participant is not informed. The question of interest often has to do with probability learning – how the participant's bias to predict one outcome or the other changes over time. The usual finding is that participants "match" rather than "maximize", meaning that over many trials their bias tends to asymptote at the level of the bias in the event generator; for example, if the "coin" is biased to produce 75% heads, then participants will, by the end of a session, predict heads on roughly 75% of trials. Under a maximizing strategy, participants would come to predict heads 100% of the time, once they detected a bias, so a matching strategy is too difficult to explain using the simplest rules of rational choice.

Probability learning studies have thus shown that people are to some extent sensitive to base rates and changes in base rates, and adjust the frequency of their predictions accordingly, if suboptimally. Base rates, though related to streaks, are a distinct source of information, with different dynamics that may make them more or less appropriate to a given decision-making scenario. Thus, a probability learning experiment involving a biased coin might track changes in the bias to predict heads as experience with the biased coin grows. We are interested in tracking changes in the bias to predict heads as a streak of heads increases in length, from one head (following a tail), to two consecutive heads, and so on. Thus, in terms of the gambler's fallacy, we are interested in how the strength of the gambler's bias might change as a function of number of losses. Similarly, in terms of the hot-hand heuristic (Burns, 2004a), in which streaks of successes serve as an adaptive allocation cue, we are interested in how the strength of the team's bias to give the ball to one shooter is affected by that shooter's recent success at scoring.

Apart from the empirical question of whether decision makers respond to streaks, there is also an important theoretical question relating streaks to base rates. Both reflect any biases toward one outcome or the other in the outcome-generating process; the base rate of that outcome will be higher, and streaks of that outcome will be longer and more frequent. However, if the bias in the outcome-generating process happens to change, as a function of a shift in environmental characteristics,

for example, the base rate will change only gradually in response. In contrast, the probability of a streak of a given length will change much more quickly. For example, a sudden increase in the heads bias of a virtual coin can immediately produce a streak of heads that is much longer than any that the decision maker experienced under the old bias; the base rate, which represents the overall historical frequency of heads, will change much more slowly. Thus, streak statistics are the more sensitive measure of change in the environment, and may play an important role in the decision making of adaptive organisms, particularly when there is variability in the bias of the outcome-generating process (Burns, 2004a).

In the next section, we describe an experiment in which participants experienced a series of coin flips and were asked after each flip to predict whether or not the next outcome would be heads. Dependent measures were (1) the overall frequency of heads predictions, and (2) the frequency of heads predictions conditional on length of heads streak (the frequency of heads predictions after one head, after two heads, and so on). We then present a model that accounts for the resulting patterns of conditional heads predictions in terms of simple memory-based decision processes; the model has available a strategy in which it asks itself whether it has seen a streak of this length before, and bases its prediction based on what is retrieved from memory. The general discussion relates our memory-based model to a formal model from the economics literature.

2. Experiment

In our experiment, each participant experienced a series of 150 coin flips, and was asked after each flip to predict the outcome of the next flip. Aiming to both replicate and generalize our results, we used two different "coins," each with a different bias towards coming up heads (60% or 75%). Participants in both groups were told that there might be a bug in the computer program generating the coin flips, such that the outcome of one flip might influence the outcome of the next; the nature of the bug was not specified, but the goal was to invite participants to view flips in causal terms, such that some underlying mechanism might be responsible for streaks of heads. Burns (2002) found this instruction to be effective in manipulating participants' beliefs about the randomness of the outcome-generating process. To draw their attention to the bias in the coin, participants were asked periodically to report a count of heads since their last report.

We also collected a measure of working memory span, given that this appears to affect what information people draw from sequences of events. Empirically, people with shorter spans can be better at detecting correlations between events (Kareev, Lieberman, & Lev, 1997). One explanation of this finding is that small samples are more likely to be extreme, and therefore are more likely than large samples to exceed a given detection threshold (for a mathematical analysis, see Kareev, 2000). For our purposes, the relevant implication is that people with small memory spans may be more likely to detect spurious correlations, so may be more likely to make inferences about sequential dependencies between events in our prediction task and therefore predict, on the basis of such perceived dependencies, that a given streak is likely to continue. We return to this link between perceived dependency and positive recency later, in context of our memory model.

3. Method

3.1. Participants

73 students recruited from the Michigan State University subject pool were randomly assigned to one of two conditions; 37 were assigned to a coin with a 75% heads bias, and 36 were assigned to a coin with a 60% heads bias.

3.2. Materials and procedure

Participants performed the working-memory task first, and then the two-choice prediction task. Both tasks were administered on-line using PCs programmed in JavaScript to control standalone web pages.

The working memory task, based on the operation span task of Kane et al. (2004), required participants to remember words while verifying the accuracy of arithmetic equations. To-be-remembered words were presented one at a time for one second each. Each word was followed by a question of the form, "Is $(6 \times 2) - 5 = 7$?". Participants mouse-clicked a "yes" button if the equation was correct or a "no" button if it was not. The equation always consisted of a parenthesized multiplication as the first term, followed by the addition or subtraction of a constant. Participants were asked to respond accurately to the equations as quickly as they could, but were not given a time limit. Responding to the equation triggered onset of the next word. Equations and words alternated until participants were asked to recall all words from the current sequence in the correct order by typing them into a text box. Sequences varied from two to six words long, with two sequences of each length, but with only one sequence of length two. Thus, the total number of words was 38.

After performing the working memory task, participants read on-line instructions describing the two-choice prediction task:

> In this experiment you will observe the result of a series of penny tosses. Before each toss, you will be asked to try to predict what you think the result of each toss will be. You can take as long as you like for this, but there is nothing to be gained by waiting. The result of each toss should be random, but the coin may be biased towards heads or tails, and there may be a bug in the program such that one flip may influence the result of the next flip. After selecting "heads" or "tails", you will see the result of the coin toss. Every so often, we will ask you some questions about the task, including how many "heads" have been flipped since we last asked. (You can't keep a written record, so you have to remember this.)

Participants then had an opportunity to ask questions, after which they predicted the outcome of a training flip. This and subsequent coin flips unfolded as computer animations each taking 3 s. Predictions for the outcome of each flip were made by clicking one of two buttons labeled "heads" and "tails".

Participants then observed 150 flips, without knowing how many flips there would be. Every 30 flips they were interrupted and asked the following four questions:

1. *How many times has the coin come up heads?*
2. *How well do you think you are doing at the task?* Responses were given using a scale from 1 (very poorly) to 7 (very well).
3. *How random do you think have been the flips produced by the program?* Responses were given using a scale from 1 (not at all random) to 7 (completely random).
4. *If you have an idea about any error in how the program generates flips, please write it below.* Responses were entered in a free-form text box.

The purpose of questions (3) and (4) was to focus participants on the possibility that some underlying mechanism could be responsible for producing streaks of heads.

3.3. Design and analyses

For the two-choice prediction task, participants were assigned to one of two groups, which differed in the heads bias of the coin (75% or 60%); the two groups were analyzed separately. The first 50 trials of each session were excluded from analysis, as is standard in probability learning experiments (e.g., Estes, 2002). Across remaining trials, we measured the overall *baseline prediction bias* towards heads, and, as described below, the *conditional prediction bias*, or the bias to predict heads as the next outcome conditional on the length of the heads streak to that point. Conditional prediction biases were submitted to analysis of variance (ANOVA) with length of heads streak as the independent variable. The length variable had levels 1 to 8 in the 75% condition and levels 1 to 5 in the 60% condition, reflecting the constraint that each experimental session had to contain at least one instance of a given length for that length to be included as a level; the 75% condition produced more long streaks, so more lengths were included.

For the working-memory task, a recalled word was scored as correct only if it was given in the correct position in the sequence. To examine the relation between working memory capacity and performance in the two-choice prediction task (conditional prediction biases, in particular), we assigned each participant to one of three groups according to their score on the working memory task. Boundaries for low, medium, and high capacity were determined from an independent sample of 498 participants, recruited from the same subject pool, to which the working memory task had been administered (from unpublished data set, Burns, 2004b). This large sample allowed us to validate the working memory task, which had similar reliability and correlations with intelligence (based on ACT scores) as other measures of verbal working memory (Kane et al., 2004). We partitioned this sample into three equal-sized groups, producing boundary scores (out of a possible high score of 38 words correct) of 21 and lower for low capacity, 22–28 for medium capacity, and 29 and higher for high capacity. Applied to our 73 participants, this partition produced group sizes in the 75% condition of 12 low, 9 medium, and 16 high capacity participants, and in the 60% condition of 9 low, 13 medium, and 14 high capacity participants.

4. Results

Fig. 1 (filled bars) shows the baseline bias to predict heads. The difference in baselines across the two groups reflects the difference in bias of the coin, with participants more likely to predict heads when the heads bias was greater.

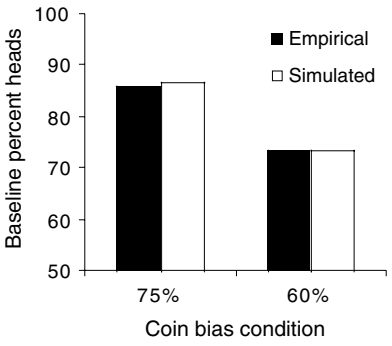

Fig. 1. Baseline bias to predict heads for the 75% and 60% coin bias conditions.

Figs. 2 and 3 (left panels) show conditional prediction biases for 75% and 60% conditions, respectively. The x-axis represent streak length: 1 means one head after a tail, 2 means two consecutive heads after a tail, etc. The y-axes represent conditional prediction biases expressed in terms of deviation from baseline bias. For example, a value of 2 for streak length 1 would indicate that the bias to predict heads after a streak of one head was 2% greater than the baseline bias to predict heads (in that coin-bias condition). Positive values thus indicate positive recency – a greater-than-baseline tendency to predict that a streak will continue – and negative values indicate negative recency.

For the 75% condition, the main effect of streak length on conditional prediction bias was reliable, $F(7,238) = 3.4$, $p < 0.03$, as was the quadratic trend, $t(238) = 2.4$, $p < 0.03$. For the 60% condition, the main effect of streak length was not reliable, $p = 0.164$, but the quadratic trend was, $t(140) = 2.6$, $p < 0.02$.

With working memory capacity added as a factor to the ANOVAs, the interactions of capacity and streak length were not reliable. Nonetheless, Fig. 4 suggests that differences in working memory capacity explain some of the variance in conditional prediction bias in the 60% condition (no clear patterns emerged in the 75% condition). The left panels show conditional prediction biases for low, medium, and high capacity groups, and the indication is that participants with smaller capacity have a greater tendency to exhibit positive recency. We return to these effects later in discussing our model, where we present a theoretical reason to believe that working memory capacity should have reliable effects when tested with sufficient statistical power.

5. A memory model

We explain the U-shaped pattern in Figs. 2 and 3 in terms of use of the historical frequency of streak lengths as a heuristic. In particular, we assume that if the decision maker is experiencing a streak, and

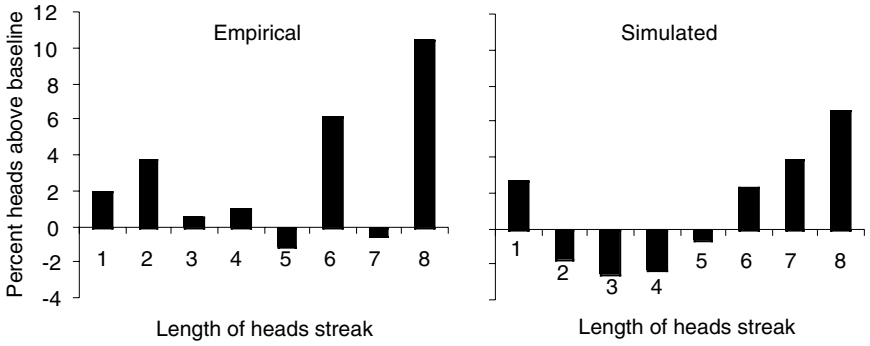

Fig. 2. Percent heads predictions above baseline for the 75% coin bias condition, conditional on length of heads streak.

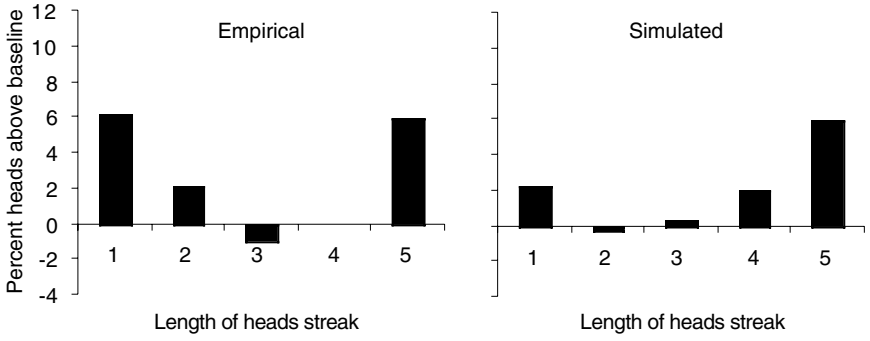

Fig. 3. Percent heads predictions above baseline for the 60% coin bias condition, conditional on length of heads streak.

can recall encountering a longer streak in the past, he or she exhibits positive recency, on the assumption that the past predicts the future. If, on the other hand, the decision maker is unable to recall any streak at all, he or she assumes that the current streak reflects a change in the environment that has made a streak of the current length more probable – and again exhibits positive recency. This logic is common in decision-making models predicated on representativeness and related heuristics. In particular, Rabin (2002) argues that decision makers "overinference" based on streaks, meaning that a streak that is unusually long in their experience leads them to overestimate the bias toward that outcome in the outcome-generating process, relative to the estimate that would be produced by pure Bayesian updating (with its infinite historical window). Viewing this overinferencing in adaptive terms, it may make sense in a changeable environment to factor a new experience into one's beliefs about the frequencies with which different outcomes are generated, and perhaps experiment with revised beliefs in future decisions.

Our model is implemented in the ACT-R cognitive theory (version 4.0, described in Anderson & Lebiere, 1998), which incorporates in a variety of ways the notion that the past predicts the future; we exploited the declarative memory mechanism, in which the activation and availability of memory elements is linked to their statistical patterns of use in a given task environment. The model fits are shown in Figs. 1–3. The empirical data are noisy, and replications currently under way will help to smooth out the curves, but the figures show that the model qualitatively captures the quadratic trends.

Our model-fitting strategy was to fit the baseline and conditional prediction biases from the 75% condition (Fig. 1, left, and Fig. 2), then to vary the smallest possible number of parameters (which

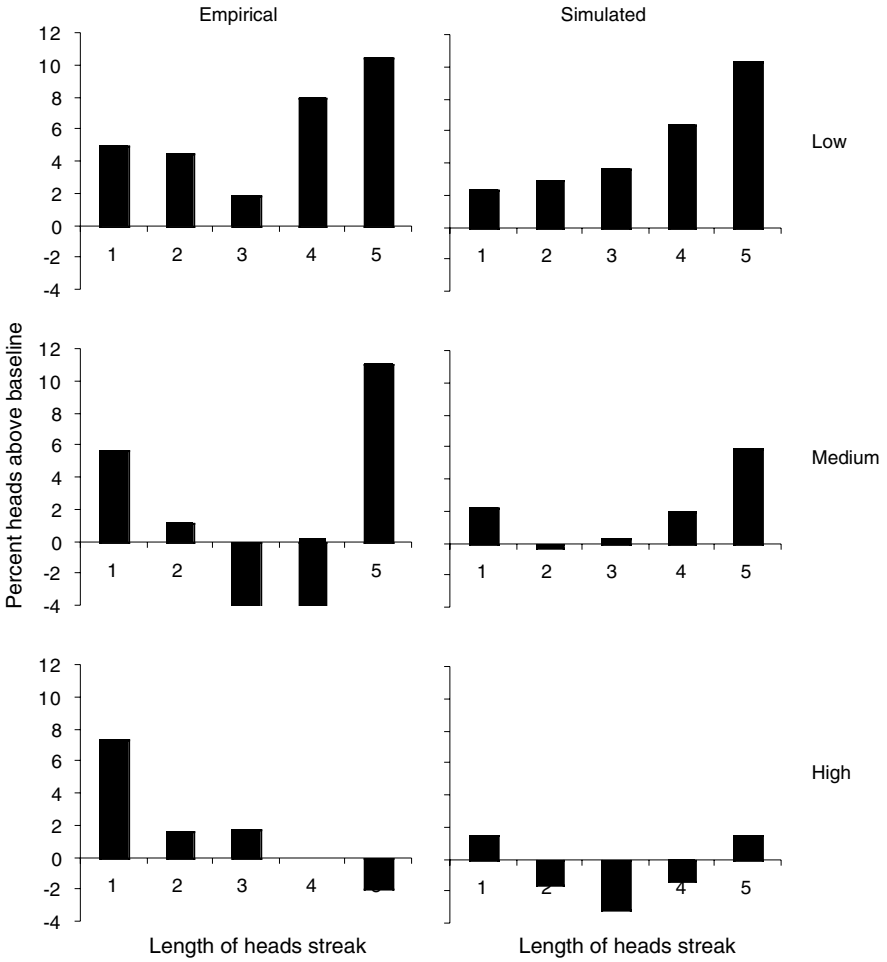

Fig. 4. Percent heads predictions above baseline for the 60% condition, conditional on length of heads streak, for low, medium, and high working memory capacity. Empirical capacities were determined by tripartite split, and simulated capacities were implemented by varying the ACT-R goal-activation parameter across values of 0.8, 1.0 (the default), and 1.2.

turned out to be one) to fit the baseline prediction bias from the 60% condition (Fig. 1, right). Encouragingly, this one parameter change concurrently produced the fit to the conditional prediction biases in the 60% condition (Fig. 3), suggesting that the model is capturing basic sources of variation in the human response to the different conditions.

6. A process-level view

Here we describe the processes that execute on individual trials, as the model observes an outcome (heads or tails) and makes a prediction for the next outcome, and link them to the U-shaped pattern in conditional prediction bias. In a later subsection, we examine the model's predictions for the effects of working memory capacity.

On most trials, the model stochastically selects one of three processes to predict the next outcome. Two of these processes are simple; one guesses heads and one guesses tails (regardless of the current outcome). Guess-heads is biased to be selected more frequently than guess-tails, reflecting the bias of the coin being flipped; the probability learning that produces this bias is of secondary interest here, and we currently do not attempt to model it.

The third process, *heuristic-predict*, is more complex, implementing the frequency-based heuristics sketched above. If this process is selected on a given trial, it spans the next several trials, exiting when the current streak terminates; while this process is active, guess-heads and guess-tails are locked out. The intuition is that the model can wave in and out of using frequency heuristics to guide its decision making, but when it waves in, it remains focused on that heuristic until the current streak ends. While active, the frequency-heuristic process is responsible for updating a count of the current streak in the model's mental focus of attention, retrieving past streak lengths from memory as guidance for predicting the next outcome, and, when the streak is over, encoding the length of the just-ended streak in episodic memory. All such streak lengths are represented as distinct items in ACT-R's declarative memory.

On a given trial, having first updated its count of the length of the current streak, the heuristic-predict process then tries to retrieve a streak length from memory; that is, it tries to retrieve any streak length encoded by heuristic-predict in the past. The retrieval is primed by the length of the current streak; thus, if the model has tracked four heads in a row, memories for streaks of length four will be more active, other things being equal, than memories for streaks of other lengths. This associative activation diminishes with integer distance from the current count, with length-four memories receiving more than length-three and length-five memories, which in turn receive more than length-two and length-six memories.

Three decision rules govern how the heuristic-predict process maps the outcome of an attempted memory retrieval to a prediction for the next coin flip. Successful retrieval is not guaranteed, but if a streak length is retrieved, the model compares the length it represents to the length of the current streak. If the retrieved length is longer than the current length, the model takes this as reason to expect the current streak to continue, so it predicts a repeat of the current outcome. If the retrieved length is the same as or shorter than the current length, the model takes this as reason to expect the current streak to end, so it predicts the opposite of the current outcome. Finally, if retrieval fails altogether, the model takes this as evidence that the current streak is unusually long, and again predicts a repeat of the current outcome. Recapping the logic, the notion is that it may make sense, in a changeable environment, for the decision maker to be open to the possibility that the rate at which an outcome is generated has changed. Here, although there is no explicit revision of beliefs about the environment, there is a decision rule that interprets a retrieval failure as evidence of a novel circumstance.

Each of the three prediction processes – guess-heads, guess-tails, and heuristic-predict – is associated with a probability of being selected on a given trial (when heuristic-predict is not already active). These three parameters were among those adjusted to fit the 75% data. Having fit those results, we held all parameters constant in fitting the 60% data, except the probability of selecting guess-tails. This was increased, reflecting greater frequency of tails outcomes given the drop to a 60% heads bias in the coin.

7. An implementation-level view

Here we briefly describe the implementation of the model, in terms of ACT-R constructs. The model code is available for downloading at http://www.msu.edu/~ema/streaks.

The simpler prediction processes – guess-heads and guess-tails – are each represented by one ACT-R production. The selection probability of each production is represented using ACT-R's production utility parameters, which are the inputs to the conflict resolution scheme in this particular production system (selection probabilities are therefore represented only indirectly, in terms of utilities).

The heuristic-predict process is represented by a set of productions operating on a variety of memory representations. The process is initiated by a *start-count* production, which competes in conflict resolution with guess-heads and guess-tails. If start-count is selected on a given trial, it enables an *increment-count* production on the following trials, and eventually a *stop-count* production, which is selected in response to the outcome that ends the streak.

Stop-count is responsible for encoding a declarative memory element representing the length of the streak when it ended. On trials on which heuristic-predict is active – that is, on trials after start-count has fired to activate the process but before stop-count has fired to terminate it – guess-heads and guess-tails are excluded from conflict resolution.

The predictive component of the heuristic-predict process begins with a *retrieve-streak* production, triggered once the increment-count production has fired on the current trial. Retrieve-streak attempts to retrieve some streak length encoded by stop-count in response to a past streak. This retrieval attempt can fail, if no potential target's activation level is above a threshold that is a system parameter. The retrieval attempt can succeed with any streak length in memory that is active enough, with the most active streak length at that instant being the one that satisfies the retrieval request. Therefore, activation dynamics, reflecting decay, priming, and transient noise, govern which streak length is retrieved. Decay, characterized by ACT-R's base-level learning mechanism, means that older streaks have less activation and thus are less likely to be retrieved; this suggests, as a test of the model, engineering particular distributions of streak lengths within a session to see if model and humans respond in similar ways.

Priming (associative activation) flows from the current streak length, stored in the system's mental focus of attention, to all streak lengths in memory, weighted by their integer distance from the current streak length. Thus, as described above, if the current streak is of length four, past streak lengths of four receive the most associative activation, past streak lengths of three and five the next most, etc. This priming gradient plays a critical role in producing the quadratic trend in conditional heads predictions. Longer streaks are less frequent than shorter ones, so as the current streak grows, the probability of retrieving any streak length from memory decreases, because fewer streak lengths receive the maximum amount of priming (which may be necessary to bring a target above threshold). Failure to retrieve a streak length triggers one of the decision rules described above, namely to predict continuation of the streak, on the assumption that the bias of the outcome-generator may have suddenly changed. If the next outcome is as predicted, the heuristic-predict process will continue to be active, and will ultimately encode a long streak in memory. Thus, the second and later instances of a long streak will be less likely to trigger this particular decision rule, given the model's prior experience.

Note that the heuristic-predict process is generic with respect to outcome, meaning that in principle another test of the model would be against the empirical biases in tails predictions.

The model performs the same 150 trials as do human subjects, so in this sense its task environment is veridical; however, we did not try to match the timing parameters of the experiment, so, for instance, there is currently no account of three-second duration of the animated coin flip. We would expect, though, that the changes required accommodate the actual time course of the experiment would be absorbed by existing model parameters.

The parameters adjusted to fit the 75% data (baseline and conditional predictions) include the utilities of guess-heads, guess-tails, and start-count; the activation threshold for retrievals (ACT-R's :rt); and two parameters (peak and slope) that determine the gradient with which priming flows from the current streak length to the same and other streak lengths in memory. To then fit the 60% data, we adjusted only the guess-tails utility; this utility was increased, consistent with the decrease in the heads bias of the coin. The ACT-R decay parameter (:bll) was set to its default value of 0.5, and the activation noise (:ans) and utility noise (:egs) parameters were both set to 0.3.

8. Effects of working memory capacity

Fig. 4 compares empirical conditional prediction bias, broken out by memory capacity, with simulation results from the model (right panels). To alter the model's effective memory capacity, we varied an ACT-R parameter (goal activation, or :ga) that previous work has linked to individual differences in working memory capacity (Lovett,

Reder, & Lebiere, 1999). (We used values of 0.8, 1.0, and 1.2 for low, medium, and high capacity respectively. The default value is 1.0, so the middle right panel in Fig. 4 simply repeats the right panel of Fig. 3.) As the model's working memory capacity decreases, its tendency to exhibit positive recency increases. Importantly, this is qualitatively consistent with the finding, mentioned earlier, that people with shorter memory spans are more likely to detect correlations between events represented in memory (Kareev et al., 1997), perhaps because smaller samples are more likely than large samples to be extreme (Kareev, 2000). In the model, the heuristic-predict decision rule is triggered in response to detection of a streak that seems unusually long; with a smaller working memory capacity, the model is less likely to retrieve a streak of similar length from memory even if such a streak length is stored there, and thus more likely to consider the current streak to be unusually long (and, by the logic of heuristic-predict, expect the streak to continue).

Visual inspection of Fig. 4 suggests that effects of working memory capacity on the model's performance are capturing some of the variance in the empirical data. To provide some quantitative support, we conducted a crude maximum-likelihood analysis, computing measures of fit (root mean squared error, or RMSE) for each variant of the model to each group of participants and asking which variant produced the best fit. For each group of participants, the model variant with the corresponding capacity level was the one that provided the best fit. For low capacity participants, RMSE for the low capacity model was 1.72; for the medium-capacity model it was 4.19, and for the high-capacity model it was 6.92. For medium capacity participants, RMSE for the medium capacity model was 5.23; for the high capacity model it was 5.32, and for the low capacity model it was 7.07. Finally, for high capacity participants, RMSE for the high capacity model was 4.05; for the medium capacity model it was 4.43, and for the low capacity model it was 6.64. Thus, at least in terms of descriptive statistics, variations in the model's working memory capacity do seem to capture some variance in the empirical data, though the amount of empirical data for each level of memory capacity was quite small. More generally, however, there is a theoretical basis for expecting that, in future studies with greater statistical power, the empirical interaction effect of working memory capacity and streak length on conditional prediction bias should prove to be reliable.

9. General discussion

Previous attempts to reconcile positive and negative recency heuristics in a single model (Rabin, 2002) have appealed to the law of small numbers. This "law" is a belief, named by analogy to the law of large numbers, that the frequency of events in a sample should match the frequency of those events in the population (Tversky & Kahneman, 1971). The law of small numbers affords one explanation of the gambler's fallacy (negative recency), in which, for instance, a series of red outcomes on a roulette wheel leads the gambler to bet on black next. The black outcome is thought to be necessary to "even out" the preceding streak of red outcomes in the current (relatively small) sample.

Although the law of small numbers is a fallacy in the context of gambling, where events are independent and sampling is with replacement, it has some validity when events are sampled without replacement, even if they are independent (Rabin, 2002). If a red ball is drawn from an urn containing red and black balls and not replaced, then the probability of the next ball being black is higher than it was before the red ball was drawn. For large urns this effect is minimal, but a decision maker applying the law of small numbers effectively sets the urn size to be quite small, such that each new outcome in a streak warrants a relatively large update in expectations concerning the next outcome. Thus, the longer a streak of reds becomes, the further the decrease in the subjective probability that the next ball will be red. Thus, the law of small numbers, applied to an outcome generator in which events are sampled from a (small) urn, offers one explanation of the decrease in positive recency across shorter streak lengths in Figs. 2 and 3.

The law of small numbers may also explain the upward trend in positive recency that follows the initial downward trend. Rabin (2002) proposes that

as a streak grows, there is not only Bayesian updating of the probability of the next ball being red or black, but also updating of beliefs about the relative proportions of reds and blacks in the urn. As more and more red balls are drawn, the decision maker comes to believe that there are more red balls in the urn than he or she previously thought. This effect should be greater for decision makers applying the law of small numbers, because the smaller the (mental) urn, the less likely the subjective probability of a streak of a given length. The smaller the urn, therefore, the more a streak indicates that the decision maker's beliefs about relative proportions of red balls and black balls in the urn need revision. Thus, as more and more red balls are drawn, the expectation that the next ball will be red should begin to increase again, at the point where the decision maker revises his or her subjective probability of reds and blacks in the urn. Rabin's is not a process model, so how these different forms of Bayesian updating may interact is unclear. Our model thus builds on his work by offering a precise formulation of how positive and negative recency interact.

Rabin (2002) draws support for his model from evidence that investors tend to under-react to a firm's financial prospects in the short term and to over-react in the long term. Empirically, stock prices tend to auto-correlate positively in the short term (a period of months), which can be interpreted to mean that investors insufficiently react to good or bad news. This is a form of positive recency, to the extent that it reflects a willingness to bet that recent trends in stock price will continue – for example, that a firm with a low stock price will continue to have a low stock price, despite a recent money-making breakthrough. On the other hand, stock prices tend to auto-correlate negatively in the medium term (a three- to five-year horizon), suggesting, for example, that such breakthroughs are factored only belatedly into investor decisions. This is a form of negative recency, to the extent that it reflects a willingness to accept that less recent trends in stock price may now reverse themselves. Qualitatively, then, there appears to be a mapping from Rabin's analysis of investor data, couched in terms of his urn model, to our U-shaped curve, which we explain in terms of memory processes. Of course, investor behavior unfolds over much longer time spans than the sequential decision used in the present study, so the underlying memory dynamics may change (Anderson, Fincham, & Douglass, 1999), and may in fact make different predictions about how decisions makers respond to streaks.

As a model, the law of small numbers by itself is problematic in that it can equally well explain positive and negative recency, as we outlined above, and thus has no predictive power (Burns & Corpus, 2004). Similarly, it seems unlikely that people really think in terms of "urns" (Rabin, 2002), although this is a representational hypothesis that remains to be tested. In our model we have tried to explain the U-shaped curve in terms of basic cognitive processes, linked ultimately to memory. A mapping could be made between urn size and memory capacity, and from different types of Bayesian updating to different reactions to different recall events. Thus, our model can be viewed as putting cognitive processing flesh on previous verbal theories. Whether its memory processes and simple decision rules generalize to account for use of recency biases in other decision-making tasks, and across longer time spans, are important questions for future research.

Finally, although any number of different models could presumably explain the curves shown in Figs. 2 and 3, we have presented preliminary evidence, in Fig. 4, linking the shapes of these curves to working memory capacity. This is converging evidence that memory is implicated in a basic way in decision makers' responses to streaks, and suggests a future line of inquiry in which memory load and related manipulations could be used to test specific quantitative predictions of the model.

Acknowledgements

We thank the reviewers for the International Conference on Cognitive Modeling for their helpful suggestions for improvement.

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of*

Experimental Psychology: Learning Memory and Cognition, 25, 1120–1136.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.

Burns, B.D. (2002). Does the law of small numbers explain the gambler's fallacy? Paper presented at the Forty-Third Annual Meeting of the Psychonomic Society, Kansas City, MO.

Burns, B. D. (2004a). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand". *Cognitive Psychology, 48*, 295–331.

Burns, B. D. (2004b). A webpage-based operation-span measure. Unpublished raw data.

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

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9*, 3–25.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*, 397–403.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*, 278–287.

Lebiere, C., Gray, R., Salvucci, D., & West, R. (2003). Choice and learning under uncertainty: A case study in baseball batting. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Lovett, M. C. (1998). Choice. In J. R. Anderson & C. Lebiere (Eds.), *Atomic components of thought* (pp. 255–296). Hillsdale, NJ: Erlbaum.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). New York: Cambridge University Press.

Olton, D. S. (1978). Characteristics of spatial memory. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 341–373). Hillsdale, NJ: Erlbaum.

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics, 117*, 775–816.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2*, 105–110.