

A hierarchical Bayesian model of the influence of run length on sequential predictions

Benjamin Scheibehenne & Bettina Studer

Psychonomic Bulletin & Review

ISSN 1069-9384

Volume 21

Number 1

Psychon Bull Rev (2014) 21:211-217

DOI 10.3758/s13423-013-0469-1

Psychonomic Bulletin & Review

VOLUME 21, NUMBER 1 ■ FEBRUARY 2014

PB&R

EDITOR

Cathleen M. Moore, *University of Iowa*

ASSOCIATE EDITORS

Edward Awh, *University of Oregon*

Sarah H. Creem-Regehr, *University of Utah*

Wim De Neys, *CNRS, France*

John Dunlosky, *Kent State University*

Ian G. Dobbins, *Washington University*

Matthew Goldrick, *Northwestern University*

Joseph B. Hopfinger, *University of North Carolina, Chapel Hill*

Mei-Ching Lien, *Oregon State University*

Bob McMurray, *University of Iowa*

J. David Smith, *University at Buffalo, The State University of New York*

Eric-Jan Wagenmakers, *University of Amsterdam*

Melvin Yap, *National University of Singapore*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1069-9384

 Springer



Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

A hierarchical Bayesian model of the influence of run length on sequential predictions

Benjamin Scheibehenne · Bettina Studer

Published online: 9 July 2013
© Psychonomic Society, Inc. 2013

Abstract Two models of how people predict the next outcome in a sequence of binary events were developed and compared on the basis of gambling data from a lab experiment using hierarchical Bayesian techniques. The results from a student sample ($N = 39$) indicated that a model that considers run length (“drift model”)—that is, how often the same event has previously occurred in a row—provided a better description of the data than did a stationary model taking only the immediately prior event into account. Both, expectation of negative and of positive recency was observed, and these tendencies mostly grew stronger with run length. For some individuals, however, the relationship was reversed, leading to a qualitative shift from expecting positive recency for short runs to expecting negative recency for long runs. Both patterns could be accounted for by the drift model but not the stationary model. The results highlight the importance of applying hierarchical analyses that provide both group- and individual-level estimates. Further extensions and applications of the approach in the context of the prediction literature are discussed.

Keywords Gamblers fallacy · Hot hand · Recency · Binary prediction task

Introduction

When making predictions, people often take previous outcomes into account. In basketball, for example, people

Electronic supplementary material The online version of this article (doi:10.3758/s13423-013-0469-1) contains supplementary material, which is available to authorized users.

B. Scheibehenne (✉) · B. Studer
Department of Psychology, University of Basel, Missionsstrasse
62a, 4055 Basel, Switzerland
e-mail: benjamin@scheibehenne.de

B. Studer
Institute of Cognitive Neuroscience, University College London,
London, UK

sometimes predict that a player’s probability to score increases if the previous throw was successful, an assumption that is referred to as the *hot-hand belief* (Gilovich, Vallone, & Tversky, 1985). In finance, buying stocks that have previously performed well and selling those with poor past returns is known as *momentum trading* (De Bondt & Thaler, 1985). The prediction that a run of the same outcome will continue is often referred to as *expectation of positive recency* (Burns, 2004). For binary sequences, positive recency yields the same predictions as a “win–stay, lose–shift” strategy, which is widespread in research on game theory (Nowak & Sigmund, 1993).

The reverse policy predicts that the next outcome will be different from the previous one; this describes the expectation of *negative recency*. A classic example of this strategy is the “gambler’s fallacy,” which refers to the tendency of roulette players to bet on red after the wheel has landed on black (Croson & Sundali, 2005; Laplace, 1820/1951).

Whether assumptions of positive or negative recency increase prediction accuracy depends on the structure of the environment. For positively autocorrelated (i.e., “patchy”) sequences, assuming positive recency is adaptive; for negatively autocorrelated (i.e., alternating) sequences, assuming negative recency is adaptive (Scheibehenne, Wilke, & Todd, 2011). Most previous studies seem to have found a tendency toward positive recency, with important moderators being the assumed data-generating process (Ayton & Fischer, 2004; Caruso, Waytz, & Epley, 2010), the type of presentation (Barron & Leider, 2010), and the experienced sequence length (Hahn & Warren, 2009).

Effect of run length

Croson and Sundali (2005) observed that real-life roulette players are particularly prone to the gambler’s fallacy when the other outcome has occurred over four times in succession. Barron and Leider (2010) reported a similar increase for run lengths of three or more using a virtual roulette wheel in a lab study. Others have found that perceived streakiness

monotonically increased up to a run length of three and then remained steady (Carlson & Shu, 2007).

A systematic influence of preceding run length may lead to interesting behavioral patterns. For example, decision makers might exhibit positive recency for short and negative recency for long runs. Jarvik (1951) reported initial empirical evidence for this co-occurrence of both positive and negative recency. When predicting pseudorandom series of two words, participants expected positive recency at a run of one, but turned toward negative recency for longer runs. In an experiment using coin tosses, Altman and Burns (2005) found a curvilinear relationship, with positive-recency bets occurring more often at both short and long run lengths. However, conclusions about the effect of run length that are based on existing research are somewhat limited, because investigations of behavior at runs longer than three or four are very rare.

Hierarchical Bayesian approach

Past research on recency effects in outcome predictions has commonly focused on group-level data. Although this increases statistical power, it conceals individual differences. In the extreme case, the averaged data are not representative of any single individual (Estes, 1956). This effect may also hold for research on recency effects. For example, McClelland and Hackenberg (1978) reported strong individual differences in how people predict gender in birth sequences. Likewise, Sundali and Croson (2006) found that roulette players at a casino were about equally divided into hot-hand players and gambler's fallacy players. Presumably, individual differences become even more important if run length is taken into account.

Hierarchical Bayesian techniques allow for capturing individual differences while preserving high statistical power by partially pooling the individual-level parameter estimates through higher-level group distributions (Carlin & Louis, 2009). This joint estimation of all parameters increases statistical power on the individual level and shrinks the leverage of possible outliers at the group level (Kruschke, 2010b). Here, we will outline how this approach can be fruitfully applied to research on recency effects in outcome predictions. As a starting point, we will define two competing models that are then tested on the basis of a choice experiment in which participants repeatedly indicated which of two outcomes would occur next and placed a bet on their decision.

One way to formally reconcile both positive and negative recency is to map them on a continuum that describes the probability π to predict the same outcome for the next event that has just been observed. Here, positive recency occurs for $\pi > .5$, whereas $\pi < .5$ indicates negative recency. When estimating recency effects irrespective of run length, a single

free parameter α ranging between 0 and 1 can be used to describe π across all t trials:

$$\pi_t = \alpha \quad (1)$$

As this yields an overall recency estimate, we refer to it as a “stationary” model. Since the model only depends on the immediately prior event, it satisfies path independence, or the Markov condition (Oskarsson, Van Boven, McClelland, & Hastie, 2009).

To incorporate the effect of run length, we extended the stationary model by one parameter β , which describes the change of π as a function f of the previous run length, denoted r (e.g., for a hypothetical sequence {R, B, B, B, R} of red and black outcomes, r is {1, 1, 2, 3, 1}):

$$\pi_t = f(\alpha + \beta \cdot [r_t - 1]). \quad (2)$$

Because for run lengths greater than one, recency does not just depend on the last outcome, but also on r , we refer to it as the *drift* model. Here, a positive β parameter indicates that people are more likely to predict a continuation of a streak, the longer it lasts; a negative β parameter indicates that the probability to switch increases with run lengths greater than one. For $\beta = 0$, the probability to stay or switch remains constant for different run lengths, and the drift model reduces to the stationary model. Note that the drift model allows for a combination of positive and negative recency that cannot be accounted for by the stationary model.

Since π describes a probability, its value must stay between 0 and 1, even for high values of r . Here, one option would be to apply a nonlinear link function; another option would be to retain a linear relationship between r and π by restricting the possible range of β contingent on α and r_{\max} (the longest occurring run). We chose the latter approach because the linearity assumption allows for an intuitive interpretation of the obtained parameter values, similarly to a linear regression. For a given α , we limited the range of β between $\delta^* \alpha$ (upper limit) and $\delta^* \alpha - \delta$ (lower limit), where δ is defined as $-1/(r_{\max} - 1)$. This restriction reflects the idea that the possible increase in positive recency is limited if π is high initially ($\alpha > .5$), and vice versa for $\alpha < .5$. Thus, consistent with theoretical assumptions, the β and α parameters in the drift model are negatively correlated a priori.

Since the drift model includes the stationary model as a special case, it will always provide a better fit to the observed data. Thus, the question of which model better accounts for peoples' prediction strategy cannot be answered on the basis of model fit alone, but must also take the models' complexity into account. Here, Bayesian techniques provide a principled way to compare the two models on the basis of the Bayes factor, which indicates the ratio of the observed evidence of one model over the other (Kass & Raftery, 1995). Bayesian techniques also yield precise estimates of the models'

parameters on the group and individual levels, including full joint posterior probability distributions providing information about the parameters' reliability and correlations.

Predictions

As was outlined above, predicting sequences of events may just depend on the immediately previous outcome, as described by the stationary model, or it may be contingent on run length, as predicted by the drift model. On the basis of the results of past research, we predict that the drift model will provide a better account of peoples' prediction strategies.

Within the drift model, the next question is in what manner would predictions change with run length. Many previous experiments found an initial tendency for positive recency. In this case, a decrease of π with run length (i.e., a negative β parameter) seems likely a priori, because as outlined above, this limits the range by which π can increase further. Thus, one might expect a change from positive to negative recency. Alternatively, an initial propensity for positive recency could be further strengthened as run length increased, which would yield a positive correlation between α and β .

Method

Participants

Fourty students at the University of Basel participated in the experiment in exchange for course credit or a book voucher worth 15 Swiss francs (15 US dollars) plus a performance-dependent bonus (0–5 Swiss francs). One participant always predicted the same outcome and was excluded from further analyses because these data yielded no variance, and hence no opportunity to model the responses.

Experimental task

Each participant completed a total of 256 prediction trials, presented as a computerized gambling game. Participants saw a wheel of fortune with two equally likely outcomes (red/black). On each round, they first decided on the color and then placed a bet on the outcome, ranging from 10 to 90 points. If participants selected the correct color, they won the betted points, otherwise they lost them. Participants were given an initial endowment of 500 points and were reimbursed dependent upon their final task earnings. No information was given about the rules or processes that generated the observed outcomes. Throughout the experiment, a vertical history bar displayed the outcomes of the ten preceding rounds.

Sequences

In order to present participants with longer runs of one color without eliciting the impression that sequences deviated from a random distribution, each color sequence was randomly drawn on the basis of the following rules: The first and third sets of 64 color items stemmed from a Bernoulli distribution with a base rate (i.e., the probability of red over black) and alternation probability of .5 each. The second and last sets of 64 items also had a base rate of .5, but the alternation probability was set to .3 (equivalent to an autocorrelation of .4). Thus, the mean alternation rate across the sequence was .4. To allow for a better comparison between individuals, only sequences containing exactly two runs of length eight (indicating the maximum run length, r_{\max}), three runs of length seven and six, five runs of length five, and no runs longer than eight were used, which approximately matched the expected run lengths in the case on hand.

Model implementation

The models were implemented in a hierarchical Bayesian framework. This required specifying the likelihood function and prior probabilities of all parameters, including the group level.¹ On each trial t , the likelihood of predicting the same outcome that had been previously observed (denoted x) was modeled as a Bernoulli distribution, with parameter π_{ti} indicating the probability to stay for each individual i :

$$p(x_{ti} | \pi_{ti}) = \pi_{ti}^{x_{ti}} \cdot (1 - \pi_{ti})^{1 - x_{ti}}. \quad (3)$$

Uninformative prior distributions were assigned to the group-level means and standard deviations of all parameters, such that all possible values of α (stationary model) and all possible parameter combinations of α and β (drift model) were equally probable on the individual level. For the drift model, α and β were drawn from a joint group-level distribution that also yielded their correlation. Both models were estimated using Monte Carlo Markov chain samplers based on the BUGS programming language, implemented in JAGS (Plummer, 2011) and called from within the R software (Version 2.14.0). To improve estimation efficiency and allow for proper aggregation of the individual parameter estimates, the group-level means were sampled from normal distributions and then rescaled to match the intended range on the individual level through probit transformations.

¹ See the [supplementary materials](#) for details and programming code.

Results

Participants took 28 min on average to complete the task ($SD = 3$ min) and predicted 52.1 % of trials ($SD = 3.8$ %) correctly. The average task earnings were 743 points ($SD = 1,192$), which translated to an average payoff of 1 Swiss franc.

Stationary model estimates

On the group level, the mean probability α to stay on the same color was .61, and 95 % of the highest posterior density (HPD₉₅) ranged from .56 to .67, indicating credible evidence for positive recency expectations across all individuals. As is shown in Fig. 1, considerable variability was observed across individuals: For the majority of participants ($n = 25$), α was credibly higher than .5 (positive recency); for five participants, α was credibly lower than .5 (negative recency); and for nine participants, α was not credibly different from .5. Bayes factor estimates obtained through the Savage–Dickey density ratio further confirmed the evidence

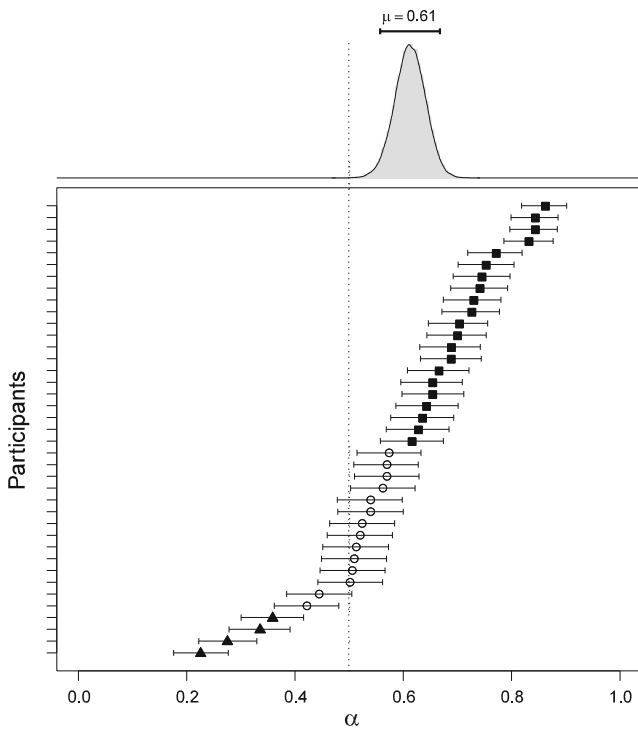


Fig. 1 Estimated posterior distribution of the α parameter in the stationary model. The distribution on top shows the posterior of the group-level mean. The single dots in the panel indicate the mean for each individual participant (sorted in decreasing order). Filled dots represent individuals for whom the evidence for positive recency (squares) or negative recency (triangles) was at least 30 times higher than the evidence for a null model assuming no recency (open circles), as indicated by the Bayes factor. Error bars indicate 95 % of the respective highest posterior density (HPD₉₅)

for positive and negative recency on the individual level (Morey, Rouder, Pratte, & Speckman, 2011).

Drift model estimates

For the drift model, the mean of the group-level α parameter was estimated at .62 (HPD₉₅ from .57 to .66). The mean β parameter estimate was 0 (HPD₉₅ from $-.03$ to $.03$), indicating that the probability of choosing the same outcome as on the last trial did not vary as a function of run length. Again, these group-level results concealed considerable individual differences. As is displayed in Fig. 2, β was credibly higher than 0 for ten of the 39 participants (26 %, positive slope), and another nine participants (23 %) were best described by a negative slope ($\beta < 0$).

Figure 3 shows the posterior means of α and β for each participant. The means were positively correlated across participants ($r = .6$, calculated from the group-level distribution), despite an a priori negative correlation. For 22 individuals (56 %), the initial tendency for either positive or negative recency was further enhanced as runs increased. The remaining 17 individuals (44 %) had $\alpha > .5$ and $\beta < 0$, indicating a decrease of positive recency as runs increased.

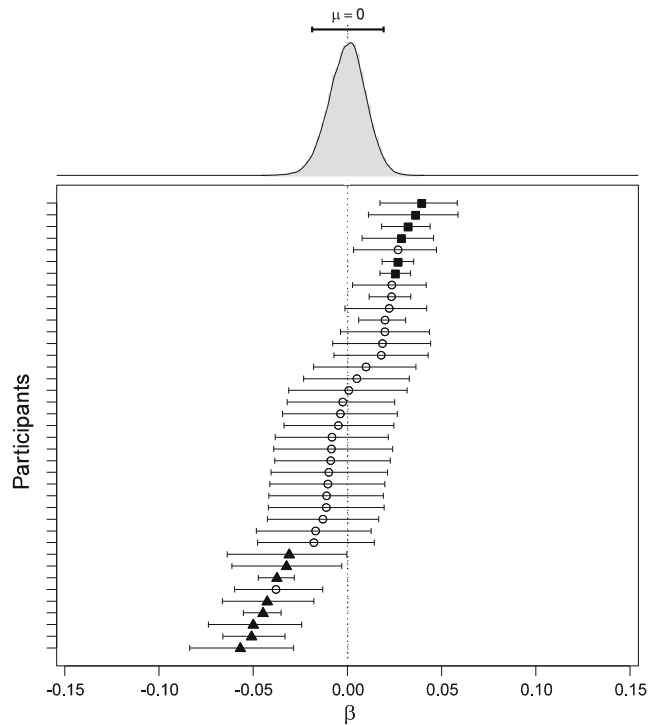


Fig. 2 Estimated posterior distribution of the β parameter in the drift model. The distribution on top shows the posterior of the group-level mean. The single dots in the panel indicate the mean for each individual participant (sorted in decreasing order). Filled dots indicate higher evidence for the drift model over the stationary model (Bayes factor > 1). Squares indicate drift toward positive recency, and triangles indicate drift toward negative recency. Error bars indicate the respective HPD₉₅

For ten individuals (26 %), this led to the co-occurrence of both positive and negative recency (for short and longer runs, respectively). The opposite case, a drift from negative recency to positive recency, did not occur.

Toward a more intuitive understanding of these parameter estimates, Fig. 4 displays the individual drift model posterior predictions of the probability to stay on the previously displayed symbol for each run length. Most lines are above .5 (indicating positive recency) and have a positive slope (indicating positive drift). A combination of both positive and negative recency occurs if the lines cross the dashed line at .5. Figure 4 also shows the observed mean probabilities of staying and the group-level prediction of the drift model, both of which are almost flat.

Model comparison

To decide whether the drift model or the stationary model is better suited to capture participants' behavior, we conducted a Bayesian model comparison on the individual level by means of the product-space method implemented in BUGS, which directly yields the Bayes factor (Carlin & Louis, 2009; Kruschke, 2010a). This comparison implicitly takes model complexity and the choice of priors into account.

The results of the comparison indicated that 64 % of participants ($n = 25$) were better described by the (nested) stationary model; the remaining participants were better described by the more complex drift model. For all but two of the participants described better by the stationary model, Bayes factor estimates were below 3, indicating weak

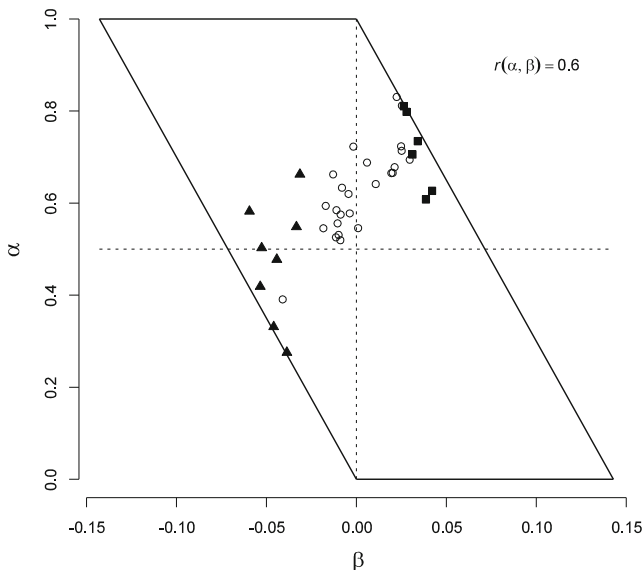


Fig. 3 Joint plot of α and β parameters. Each dot represents one individual. Filled dots indicate higher evidence for the drift model over the stationary model (Bayes factor > 1). Squares indicate drift toward positive recency, and triangles indicate drift toward negative recency. The rhomboid indicates the possible parameter space

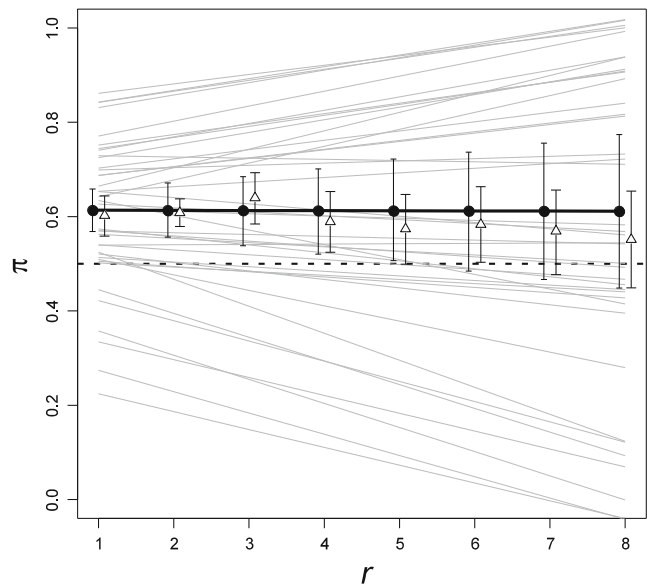


Fig. 4 Probability π to stay on the previously displayed symbol, contingent on run length r . Thin lines indicate the drift model predictions for individual participants, the bold black line indicates the group-level predictions, and the open triangles indicate the observed group-level means. Lines above the dashed line at .5 indicate positive recency, lines below indicate negative recency, and lines crossing the dashed line indicate a combination of positive and negative recency within the same individual. The slope of each of the lines indicates the direction of drift (developing upward = positive drift, developing downward = negative drift). Error bars show HPD₉₅ for the predictions and bootstrapped 95 % confidence intervals for the observed data

evidence (Kass & Raftery, 1995). In contrast, the evidence for those identified as drift model users was stronger, as indicated by Bayes factors above 3 for most of these participants. For seven participants, the Bayes factor exceeded 10, indicating strong evidence (Kass & Raftery, 1995).

Even though most individuals were described better by the stationary model, the drift model nevertheless provides a better description of the data at the group level, as indicated by the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). For the drift model, the DIC equals 1,453, which is smaller (i.e., better) than the stationary model DIC of 1,467. This seems plausible, given that run length had a strong and systematic influence for some participants.

Betting

The average bet placed per round was 59 points, with considerable variance between participants ($SD = 22$) and trials ($SD = 19$). To test whether people were more confident of bets placed on the continuation of a run than on its ending, we calculated the difference between the average betting amounts, after rescaling the bets of each individual ($M = 0$, $SD = 1$) to control for differences in absolute magnitude. On the group level, the mean difference was .2 (95 % confidence

interval [CI₉₅] from .1 to .29), indicating that on average, participants' confidence was higher when betting on the continuation of a run (i.e., positive recency).

Betting behavior further highlights the importance of run length as predicted by the drift model. On average, bet amounts were positively correlated with run length (mean $r = .07$, CI₉₅ from .05 to .1). This correlation was higher for individuals characterized by a steeper slope ($|\beta|$) in the drift model ($r = .35$, CI₉₅ from .02 to .59). Thus, people who were more likely to change their prediction policy contingent on run length did so with higher confidence as run lengths increased.

Discussion

We assessed participants' choices in a prediction task with binary outcomes and compared two models of prediction strategies using hierarchical Bayesian techniques. Two important findings were obtained. First, prediction behavior was subject to considerable individual differences and included strategies that were not apparent in the averaged group data. Second, run length influenced the predictions of some, but not all, participants.

The group-level results of both models showed that predictions were characterized by positive recency—that is, the tendency to choose the outcome that had won on the last trial. This result is in line with previous work on outcome predictions (Scheibehenne, Wilke, & Todd, 2011) and the perception of random sequences (Falk & Konold, 1997), and it seems plausible, given that on average the presented sequences were in fact slightly streaky. However, on the individual level, negative recency and indifference to previous outcomes were also observed. Similarly, although no significant influence of run length was observed on the group level, individual-level parameter estimates revealed that about a third of the participants did take run length into account. These participants were described better by the drift model than by the stationary model. Furthermore, several different prediction strategies were observed for these participants. Mostly, an initial propensity for either positive or negative recency became stronger as runs increased (see also Barron & Leider, 2010, and Croson & Sundali, 2005), and this pattern was accompanied by higher bets for long runs. In another subgroup of participants, an initial propensity for positive recency decreased with run length, leading to the co-occurrence of both positive and negative recencies, a pattern previously observed at the group level by Jarvik (1951).

In conclusion, our analyses revealed that prediction strategies varied across participants. Such variance has previously been found in a real-life investigation of roulette play (Sundali & Croson, 2006), but laboratory studies on

sequential effects in outcome predictions have largely focused on group-averaged data. Our results clearly highlight the importance of considering individual-level data, and suggest that some of the inconsistency about the nature of run length effects in previous studies might be explained by variance at the individual level.

Advantage of Bayesian techniques

Our results demonstrate that the application of hierarchical Bayesian model techniques provides a fruitful framework to quantify and test peoples' prediction strategies by taking both individual- and group-level data into account. The approach yielded exact and easily interpretable estimates for all model parameters and allowed for a rigorous model comparison taking model complexity into account (Lee, 2011; Scheibehenne, Rieskamp, & Wagenmakers, 2013). This advantage might become even more important when testing possible extensions of the drift model in future research that would allow for more complex functional relationships between run length and recency, such as curvilinear trends, step functions, or concave trajectories (Altmann & Burns, 2005; Carlson & Shu, 2007). Likewise, the ability to account for individual differences would allow for testing various differential influences on prediction strategies, such as age (Castel, Drolet Rossi, & McGillivray, 2012) or personality (Friedland, 1998).

The online supplementary material can be accessed here: <http://scheibehenne.de/OnlineSupplementScheibehenneStuder.zip>

Author Note This work was supported by Research Grant No. 100014_130149 from the Swiss National Science Foundation to the first author.

References

- Altmann, E. M., & Burns, B. D. (2005). Streak biases in decision making: Data and a memory model. *Cognitive Systems Research*, 6, 5–16.
- Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, 32, 1369–1378. doi:10.3758/BF03206327
- Barron, G., & Leider, S. (2010). The role of experience in the gambler's fallacy. *Behavioral Decision Making*, 23, 117–129.
- Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand. *Cognitive Psychology*, 48, 295–331.
- Carlin, B. P., & Louis, T. A. (2009). *Bayesian methods for data analysis*. Broken Sound Parkway, NW: Chapman and Hall/CRC.
- Carlson, K., & Shu, S. (2007). The rule of three: How the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes*, 204, 113–121.
- Caruso, E. M., Waytz, A., & Epley, N. (2010). The intentional mind and the hot hand: Perceiving intentions makes streaks seem likely to continue. *Cognition*, 116, 149–153.

- Castel, A. D., Drolet Rossi, A., & McGillivray, S. (2012). Beliefs about the “hot hand” in basketball across the adult life span. *Psychology and Aging, 27*, 601–605. doi:10.1037/a0026991
- Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty, 30*, 195–209.
- De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance, 40*, 793–805. doi:10.1111/j.1540-6261.1985.tb05004.x
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*, 134–140.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301–318.
- Friedland, N. (1998). Games of luck and games of chance: The effect of luck- versus chance-orientation on gambling decisions. *Journal of Behavioral Decision Making, 11*, 161–179.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.
- Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116*, 454–461.
- Jarvik, M. E. (1951). Probability learning and negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology, 41*, 191–297.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795. doi:10.1080/01621459.1995.10476572
- Kruschke, J. K. (2010a). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press.
- Kruschke, J. K. (2010b). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Science, 14*, 293–300.
- Laplace, P. (1951). *Philosophical essays on probabilities* (Trans: Truscott, F. W. & Emory, F. L.). Dover, New York, NY. (Original work published 1820)
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*, 1–7. doi:10.1016/j.jmp.2010.08.013
- McClelland, G. H., & Hackenberg, B. H. (1978). Subjective probabilities for sex of next child: U.S. college students and Philippine villagers. *Journal of Population, 1*, 132–147.
- Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology, 55*, 368–378. doi:10.1016/j.jmp.2011.06.004
- Nowak, M., & Sigmund, K. (1993). A strategy of win–stay, lose–shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature, 364*, 56–58.
- Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin, 135*, 262–285. doi:10.1037/a0014821
- Plummer, M. (2011). JAGS Version 3.1.0. Retrieved from <http://mcmc-jags.sourceforge.net/>
- Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review, 120*, 39–64. doi:10.1037/a0030777
- Scheibehenne, B., Wilke, A., & Todd, P. M. (2011). Expectations of clumpy resources influence predictions of sequential events. *Evolution and Human Behavior, 32*, 326–333.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, 64*, 583–639.
- Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making, 1*, 1–12.