

# Probability matching in risky choice: The interplay of feedback and strategy availability

Ben R. Newell · Derek J. Koehler · Greta James ·  
Tim Rakow · Don van Ravenzwaaij

Published online: 8 November 2012  
© Psychonomic Society, Inc. 2012

**Abstract** Probability matching in sequential decision making is a striking violation of rational choice that has been observed in hundreds of experiments. Recent studies have demonstrated that matching persists even in described tasks in which all the information required for identifying a superior alternative strategy—*maximizing*—is present before the first choice is made. These studies have also indicated that maximizing increases when (1) the asymmetry in the availability of matching and maximizing strategies is reduced and (2) normatively irrelevant outcome feedback is provided. In the two experiments reported here, we examined the joint influences of these factors, revealing that strategy availability and outcome feedback operate on different time courses. Both behavioral and modeling results showed that while availability of the maximizing strategy increases the choice of maximizing early during the task, feedback appears to act more slowly to erode misconceptions about the task and to reinforce optimal responding. The results illuminate the interplay between “top-down” identification of choice strategies and “bottom-up” discovery of those strategies via feedback.

**Keywords** Probability matching · Maximizing · Decision making · Heuristics · Rational choice theory

When faced with a choice between two options, one of which offers a higher probability of receiving a fixed payoff than

does the other, a rational agent should always choose the option with the higher payoff probability. For example, if one option delivers one dollar 70 % of the time (and nothing the rest of the time), while the other pays one dollar only 30 % of the time (and otherwise nothing), a rational agent should choose the 70 % option. This is true whether the choice is faced once or repeatedly: The option offering the higher probability of receiving the payoff should always be chosen.

Despite the clear superiority of this strategy (referred to as *maximizing*), participants faced with a series of such choices often show responding closer to *probability matching*—allocating responses to the two options in proportion to their relative probabilities of occurrence. In other words, people bet on the 30 % option 30 % of the time, and the 70 % option 70 % of the time (James & Koehler, 2011).

This striking violation of rational choice has been studied extensively over the last 60 years (Vulkan, 2000). Most probability-matching experiments have used paradigms in which participants had to learn, over successive trials, the contingencies associated with each option (e.g., Shanks, Tunney, & McCarthy, 2002).<sup>1</sup> However, some recent studies

---

B. R. Newell (✉) · D. van Ravenzwaaij  
School of Psychology, University of New South Wales,  
Sydney 2052, Australia  
e-mail: ben.newell@unsw.edu.au

D. J. Koehler · G. James  
Department of Psychology, University of Waterloo,  
Waterloo, Ontario, Canada

T. Rakow  
Department of Psychology, University of Essex,  
Colchester, UK

---

<sup>1</sup> It is important to note that in the tasks that we considered, the response options were *negatively correlated*: On any given trial, if one option delivered a reward, the other did not. This contrasts with the preparation used in many nonhuman animal-learning studies, in which the options have been *independent* (i.e., rewards can be delivered via either, both, or neither option on any trial). In such situations, nonhuman animals are often observed to allocate responses in a manner that equates the reinforcement rates per unit of consumption across choice options. This behavior is captured by the “matching law” (Herrnstein, 1997), which refers to an animal matching the relative frequency of choosing an option with the relative frequency of reinforcement (reward delivery). As was noted by Shanks et al. (Shanks et al. 2002), the term “matching law” is distinctly confusing, because it does not predict *probability matching* in the kinds of task that we consider. Instead, because the maximizing response always has the highest momentary reinforcement rate in our task, the matching law predicts maximizing at asymptote in the kind of task that we used. (See Newell, Lagnado, & Shanks, 2007, chap. 11, and Shanks et al., 2002, for further discussion.)

have investigated probability matching in tasks in which the outcomes and their probabilities of occurrence are fully described to participants (e.g., Gal & Baron, 1996; James & Koehler, 2011; Koehler & James, 2009, 2010; Newell & Rakow, 2007; West & Stanovich, 2003). The finding that probability matching is common even in these situations is remarkable, given that the described problems provide all of the information necessary for rational responding (i.e., identification of the maximizing strategy as optimal), even before a single choice is made.

Consider, for example, predicting the outcome of rolls of a ten-sided die with seven green sides and three red sides, with a fixed payment for each correct prediction. Here there is no ambiguity about the optimal strategy (always predict green), no need for a period of experimentation or exploration of the environment to determine which option is best, and no need to consider the possibility that outcome probabilities will change across successive trials. Thus, from a normative perspective (i.e., a rational economic analysis of the problem), the decision maker simply needs to identify the optimal strategy (predict green) and execute it on every trial.

Newell and Rakow (2007) examined a problem like the one described above, in which participants predicted the outcomes of 300 rolls of a simulated die. Their experiment confirmed that some participants adopted a matching strategy from the outset. However, the experiment also highlighted a further intriguing finding: Those participants who saw the outcome of each roll after making a prediction (feedback condition) demonstrated an increasing rate of maximizing across the 300 trials. By contrast, those who received no feedback continued to match, or slightly overmatch, the objective probabilities throughout the entire sequence of trials. The facilitative effect of feedback was observed despite the feedback being completely uninformative, and therefore normatively irrelevant—the die was fair, the probabilities were stationary, and the relevant outcome probabilities were already precisely known before any feedback was received. The finding that feedback nonetheless influenced the rate of maximizing is puzzling, because it suggests that people only gradually “learn” to choose optimally in a task that (normatively) requires no learning.

The finding is only puzzling, however, if one presupposes that the optimal maximizing strategy is readily generated by the decision maker in response to the initial description of the choice problem. At least for some people, the maximizing strategy might not come immediately to mind. Koehler and James (2010) gave participants a version of the die problem, but for one group they also provided a “hint” about the possible choice strategies that might be used. Specifically, before making their choices, the matching and maximizing strategies were described and participants were asked which strategy was likely to earn more money. Those given the hint (question) subsequently made

significantly more maximizing choices than those who had not been. This result suggests that the hint increased the “availability” of the maximizing strategy and implies that maximizing does not necessarily come to mind readily when the fully described problem is presented. In short, even in a described problem in which participants have all of the information needed to identify the maximizing strategy as superior, that strategy may simply fail to come to mind. The observed impact of the hint manipulation implies that probability matching comes to mind more readily than maximizing as a candidate choice strategy in response to the problem description.

One possible interpretation of the facilitative effect of outcome feedback in a fully described problem is that feedback assists “top-down” identification of maximizing as the superior strategy. That is, outcome feedback may encourage monitoring of the reward rate for the currently used strategy as well as a search for alternative strategies that might increase payoffs. In the process, some participants who initially engage in matching may “discover” the superior maximizing strategy over the course of the choice sequence. Indeed, a number of researchers have recently attempted to characterize the processes by which participants use outcome feedback to search for an optimal strategy in the binary prediction task (e.g., Gaissmaier & Schooler, 2008; Otto, Taylor, & Markman, 2011; Shanks et al., 2002). According to this interpretation, providing outcome feedback and providing a “hint” in a described problem both increase maximizing behavior by making that strategy more available than it would otherwise be (cf. Biele, Rieskamp, & Gonzalez, 2009).

To test this hypothesis, we ran two experiments using a version of the die problem in which we crossed, factorially, the presence/absence of the hint used previously by Koehler and James (2010) and the trial-by-trial outcome feedback used by Newell and Rakow (2007). If the impact of feedback was to spur the generation of alternative strategies, which eventually led to identification of maximizing as the optimal response, then we should see an interaction whereby little or no effect of feedback would be found among those participants given the hint (for whom the maximizing strategy was already highly available). This should be especially true of the subset of participants who, when prompted with that hint, correctly identified the maximizing strategy as superior.

An alternative possibility was that the hint effect observed by Koehler and James (2010) would not generalize to tasks involving trial-by-trial predictions with feedback. In their studies, the participants only made predictions for ten rolls, and did so without any information about the outcomes. In a feedback version, it is possible that seeing the regular occurrence of the low-probability outcome might make it difficult to stick to the maximizing strategy, even if, when prompted, one could correctly identify it as optimal (cf. Goodnow, 1955).

The contribution of our new experiments was that they allowed us to examine the joint influences of two manipulations—feedback and provision of a hint—that up to now have only been studied in isolation. Investigating how these two manipulations interact and the time course of their effects, through both examination of the behavioral results and application of reinforcement-learning models developed for similar tasks (Biele et al., 2009; Yechiam & Bussemeyer, 2005), would help shed further light on the cognitive underpinnings of probability matching, and maximizing, in the binary prediction task.

## Overview of the experiments

In both experiments, participants played a “die game” in which payment was contingent on making correct predictions about the outcome of rolls of a ten-sided die with seven green and three red sides. The independent variables in both experiments were the provision of a hint (present/absent) preceding the prediction task (die game), intended to make both the maximizing and matching strategies readily available, and of feedback (present/absent) regarding the outcome of each roll during the prediction task itself. The principal differences between Experiments 1 and 2 were the use of a real (Exp. 1) or a virtual (Exp. 2) die and the number of prediction trials (50 and 300 in Exps. 1 and 2, respectively). The real die was used in Experiment 1 to counter the criticism that participants might make suboptimal predictions in virtual tasks because they doubt that the die is fair and that the probabilities are stationary (e.g., Hertwig & Ortmann, 2001). In Experiment 1, participants saw the same physical die being rolled on every trial, thus presumably eliminating, or at least substantially reducing, any skepticism about the task parameters.

## Experiment 1

### Method

**Participants** A total of 123 undergraduate students from the University of Waterloo participated in return for course credit and earnings from the experiment. They were randomly assigned to one of the four experimental conditions (with  $n$  per condition varying from 29 to 32). The data from eight additional participants were excluded from the analysis: five who had completed a similar task before, two who failed to follow the instructions, and one who was red–green colorblind.

**Procedure** Participants predicted the outcomes for a series of 50 rolls of a ten-sided die with seven green sides and

three red sides. The participants were told that for every correctly predicted outcome they would receive \$0.10 (for reference, 1 CAD = ~1.0 USD). Predictions were made on a sheet of paper that organized the 50 rolls into five “games” consisting of ten rolls each. The experimenter, who was present when the predictions were made, rolled the die to determine the outcomes. This was done either after each prediction (feedback condition) or after all 50 predictions had been made (no feedback condition). The outcomes were recorded directly on the prediction sheet, and the total winnings were calculated and paid at the end of the session.

Before completing the prediction task, one group of participants (hint condition) read a description of two possible strategies that could be used in the game. Specifically, they were told: “Consider these two strategies that could be used in a 10 roll game: (a) you could predict green for all 10 throws, or (b) you could predict green for 7 throws and red for 3 throws. Which strategy do you think will win more money?” Participants in the no-hint condition were not presented the alternative strategies before completing the prediction task.<sup>2</sup>

## Results

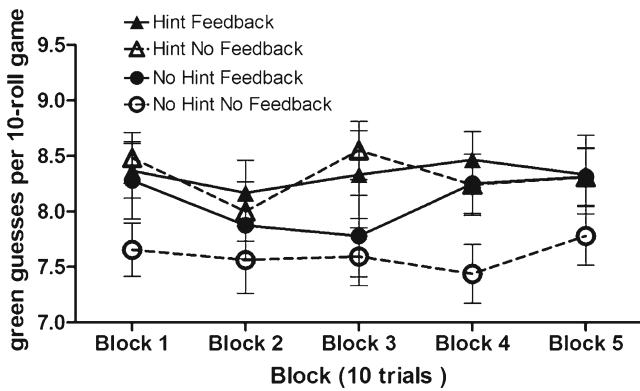
The number of times that each participant predicted the dominant color (green) in each block of ten trials was subjected to a 2 (feedback)  $\times$  2 (hint)  $\times$  5 (block) mixed-model analysis of variance (ANOVA; see Fig. 1).

We found a significant main effect of hint,  $F(1, 119) = 4.49$ ,  $p = .036$ ,  $\eta_p^2 = .036$ ; those who received the hint subsequently predicted green more often ( $M = 8.3$  per ten-roll game) than did those who did not receive the hint ( $M = 7.9$  per ten-roll game).

There was a slight tendency for predicting the dominant color more often in the feedback than in the no-feedback condition ( $M$  of 8.2 vs. 8.0 per ten-game roll). However, this effect was not statistically significant ( $p = .26$ ,  $\eta_p^2 = .011$ ). The hint-by-feedback interaction also was not statistically significant ( $p = .29$ ,  $\eta_p^2 = .010$ ). No significant effect of block emerged, nor did that factor interact with either of the between-subjects variables ( $F$ s  $< 1$ ). The mean numbers of green guesses in the first ( $M = 8.19$ ) and final ( $M = 8.18$ ) blocks were virtually identical.

Among the hint condition participants, who were asked prior to the prediction task to evaluate which strategy had higher expected earnings, 64 % correctly identified the maximizing strategy as being superior. The proportions choosing maximizing were virtually identical in the hint

<sup>2</sup> Following the prediction task, all participants in both experiments were given a questionnaire that asked about the strategies that they had adopted in the game (see Koehler & James, 2010), but those data are not considered further here.



**Fig. 1** Mean numbers of “green” guesses for a die with seven green sides and three red sides, for each of the five blocks of ten rolls in Experiment 1. Error bars indicate SEMs

no-feedback (65.5 %) and hint feedback (63.3 %) conditions. The upper panels of Fig. 2 show the numbers of green predictions by hint condition participants in the first ten-roll game, as a function of which strategy they had endorsed. The figures show that not all participants followed the strategy that they had just endorsed when making their trial-by-trial predictions. However, a comparison of the upper two (hint) and lower two (no hint) panels shows that a much higher percentage of participants endorsed maximizing in the hint condition (black bars, upper panels) than spontaneously chose green for all ten rolls of the first game in the no-

hint condition (far right open bars, lower panels). This pattern is consistent with the idea that the hint prompts participants to realize that maximizing is optimal.

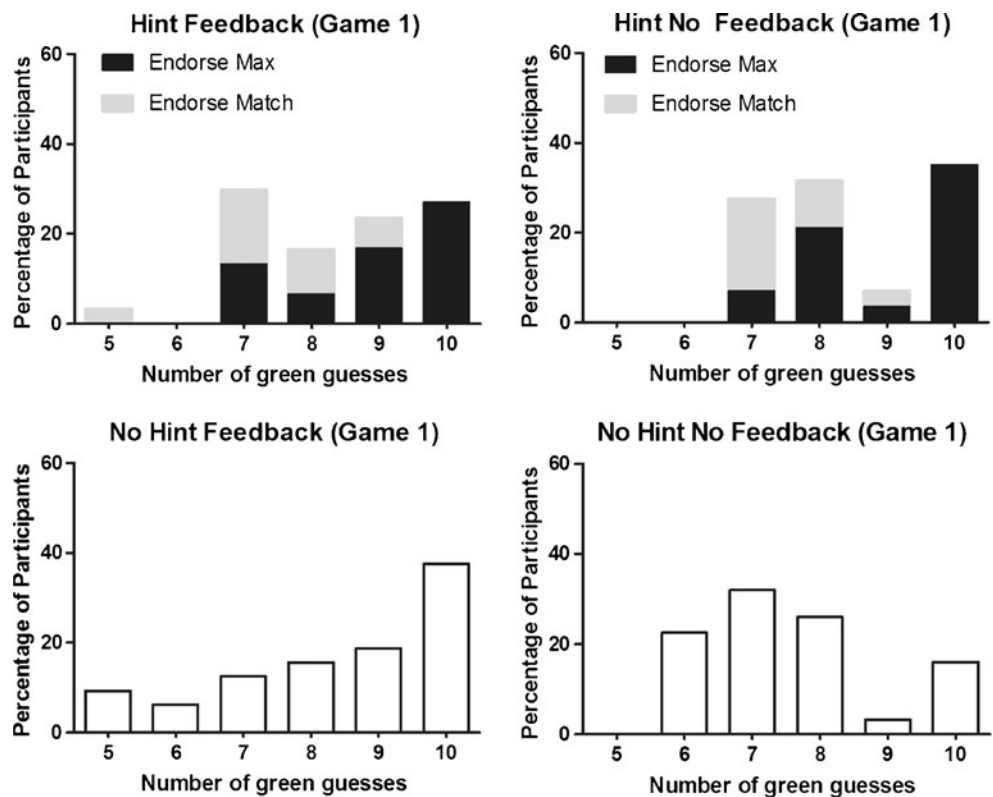
The mean number of green predictions across all trials for participants who identified maximizing (matching) as superior was 8.6 (7.8) in the hint feedback condition and 8.8 (7.4) in the hint no-feedback condition. We found no significant difference in the numbers of green guesses as a function of either block or feedback for the participants who correctly endorsed maximizing ( $F_s < 1$ ).

Discussion

The key finding of Experiment 1 is that providing a hint designed to make the maximizing and matching strategies equally available increased levels of maximizing on the subsequent prediction task. In contrast, provision of feedback about the outcome of each die roll did not significantly increase maximizing. The facilitative effect of the hint is consistent with previous findings (Koehler & James, 2010) and confirms that the effect of strategy availability generalizes to situations involving trial-by-trial feedback.

In contrast to the hint effect, we found almost no evidence that participants “learned” from feedback when no hint was provided. This result is inconsistent with previous research showing clear increases in maximizing over feedback-reinforced trials of the die game (Newell & Rakow, 2007).

**Fig. 2** Numbers of “green” guesses made by participants in the first ten-roll game of Experiment 1. “Endorse Max” and “Endorse Match” refer to the strategies selected when participants were prompted with the hint (see text). (Note that “5” on the x-axis indicates five or fewer “green” guesses.)





One difference, however, between Experiment 1 and the previous research was the relatively small number of prediction trials. Newell and Rakow found that significant increases in the rates of maximization only occurred after 100 or more prediction trials. Thus, Experiment 1 may have provided insufficient experience for “bottom-up” feedback to spur the generation of the superior maximizing strategy.<sup>3</sup> In Experiment 2, we examined this possibility by retaining the basic design of Experiment 1 but increasing the number of prediction trials to 300 (as per Newell & Rakow, 2007). This allowed us not only to attempt to replicate the findings of Experiment 1 and of Newell and Rakow, but also to examine the interplay of the hint and feedback over an extended number of predictions. The increased number of training trials also facilitated the application of reinforcement learning models to our data. These were applied in an effort to find converging evidence of the roles played by the hint and feedback in the binary prediction task.

**Experiment 2**

Experiment 2 used the same 2 (feedback: present/absent) × 2 (hint: present/absent) design, task, and procedure (except where specified below), but increased the number of prediction trials to 300 and instantiated the task on a computer to speed up the die “rolling” process.

**Method**

*Participants* A group of 100 first-year undergraduate students (28 male, 72 female;  $M_{age} = 19.8$ ,  $SD = 3.3$ ) from the University of New South Wales participated in return for course credit and performance-related remuneration.

*Procedure* The participants were told that they would play 30 games of ten rolls each (for a total of 300 trials) and would be paid \$0.02 (1 AUD = ~1.0 USD) for each correct prediction. In the two feedback-present conditions, an image of a ten-sided die rolled across the screen following each prediction and the outcome was displayed. In the two feedback-absent conditions, no die was shown and the participants were told that the die would be rolled once all of the predictions had been made, in order to determine payment. In all conditions, a participant-controlled interval followed each ten-roll game, and each game was preceded

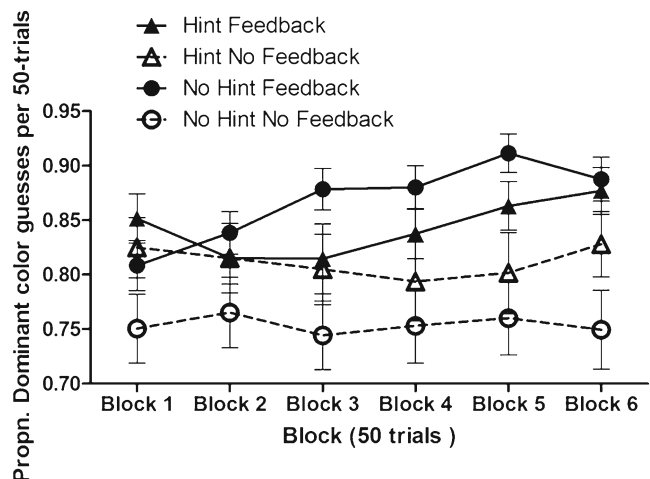
<sup>3</sup> We use the term “bottom-up,” when applied to the influence of feedback, to contrast with the perceived “top-down” influence of an intentionally applied strategy (e.g., matching or maximizing). In the present context, we imply that this “bottom-up” influence is gradual, emerging slowly over the course of successive prediction trials. We acknowledge that in other contexts “bottom-up” processes can be very rapid (e.g., perception).

by an instruction reminding participants that the same virtual die was used in every game and that they would be paid \$0.02 for every correct prediction. The dominant color (i.e., which color covered seven sides of the die—red or green) was counterbalanced across participants.

**Results and discussion**

*Behavioral results* Figure 3 plots the proportions of dominant-color predictions for the four groups of participants, averaged across the six 50-trial blocks (five “games” per block). A 2 (feedback) × 2 (hint) × 6 (block) mixed-model ANOVA revealed a main effect of feedback,  $F(1, 96) = 8.82$ ,  $p = .005$ ,  $\eta_p^2 = .079$ , a linear trend for block,  $F(1, 96) = 14.80$ ,  $p < .001$ ,  $\eta_p^2 = .134$ , no main effect of hint ( $p = .511$ ,  $\eta_p^2 = .005$ ), but a significant linear interaction between Block and feedback,  $F(1, 96) = 18.22$ ,  $p < .001$ ,  $\eta_p^2 = .16$ . The two-way interaction between hint and feedback was not significant ( $p = .11$ ,  $\eta_p^2 = .027$ ), and neither was the linear interaction between block and hint ( $p = .12$ ,  $\eta_p^2 = .025$ ) or the three-way interaction ( $p = .18$ ,  $\eta_p^2 = .018$ ).

To provide a like-for-like comparison between Experiments 1 and 2, we analyzed the initial 50 trials of Experiment 2 (i.e., the number of trials used in Exp. 1). Therefore, for Block 1 alone, a 2 (hint) × 2 (feedback) ANOVA was used to examine the proportion of dominant-color guesses. This ANOVA revealed that those given the hint made significantly more dominant-color guesses ( $M = .83$ ) than did those not given the hint ( $M = .77$ ),  $F(1, 96) = 5.06$ ,  $p = .027$ ,  $\eta_p^2 = .050$ , but there was no significant effect of feedback ( $p = .11$ ,  $\eta_p^2 = .027$ ). This replicates the results of Experiment 1 (for the equivalent number of trials). It shows that the change from a real to a virtual die, the change in population (Canadian vs. Australian participants), and the absence of the experimenter in Experiment 2 did not significantly alter the results. As a



**Fig. 3** Mean proportions of dominant-color guesses for each of the six blocks of 50 rolls in Experiment 2. Error bars indicate SEMs

complementary analysis, the Block6 data were analyzed using the same ANOVA design. This ANOVA on Block6 revealed essentially the reverse pattern of significances to that found in Block 1. By Block6, the hint no longer had a significant effect ( $p = .22$ ,  $\eta_p^2 = .016$ ), whereas the proportion of dominant-color predictions was significantly higher with feedback ( $M = .87$ ) than without ( $M = .78$ ),  $F(1, 96) = 11.78$ ,  $p = .001$ ,  $\eta_p^2 = .109$ . Neither of these individual block analyses showed significant interactions between hint and feedback.

In short, the hint manipulation had an influence on early prediction trials but not on later ones, leading to no overall main effect of hint. By contrast, the feedback manipulation had no detectable influence on early prediction trials but did exert an influence on later trials, yielding an overall main effect of feedback as well as a Feedback $\times$ Block interaction.

Of those participants who were asked prior to the prediction task to evaluate which strategy had higher expected earnings, 74 % correctly identified the maximizing strategy as being superior. The proportion choosing maximizing was higher in the hint no-feedback (21/25 = 84 %) than in the hint feedback (17/26 = 65 %) condition, but this difference was not statistically significant,  $\chi^2(1) = 2.32$ ,  $p > .05$ .

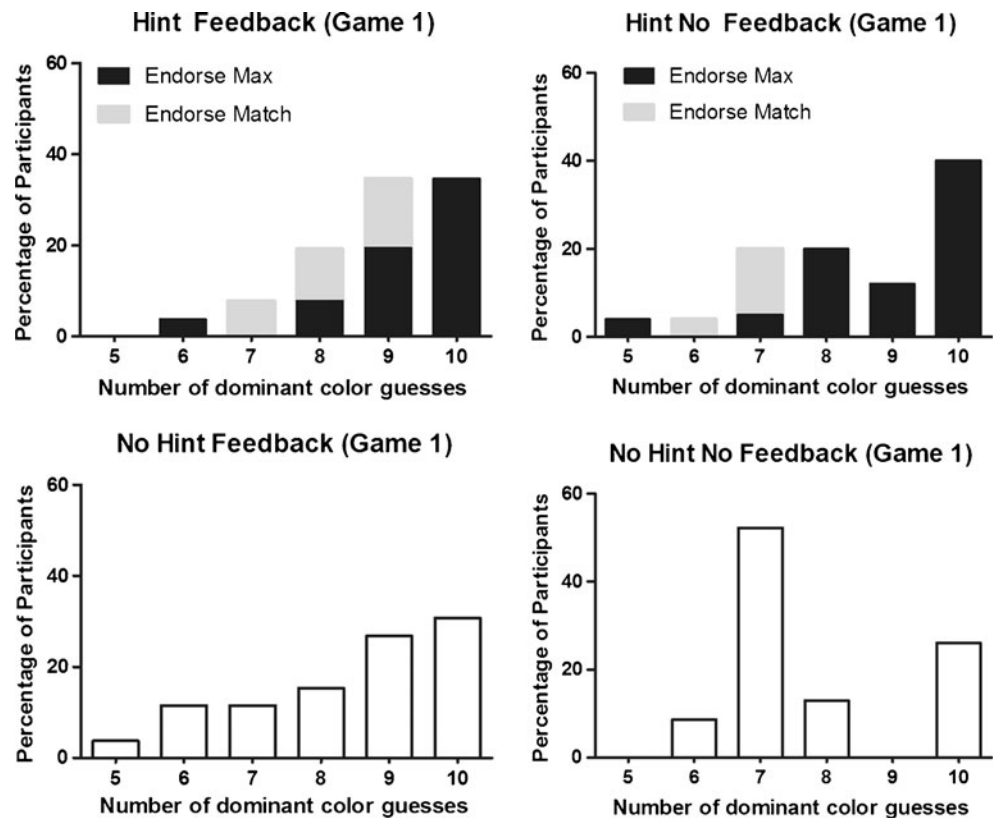
The upper panels of Fig. 4 plot the numbers of dominant-color guesses made in the first ten-roll game by participants who endorsed either maximizing or matching in the hint question. Consistent with Experiment 1, although not all participants behaved strictly according to the strategy that

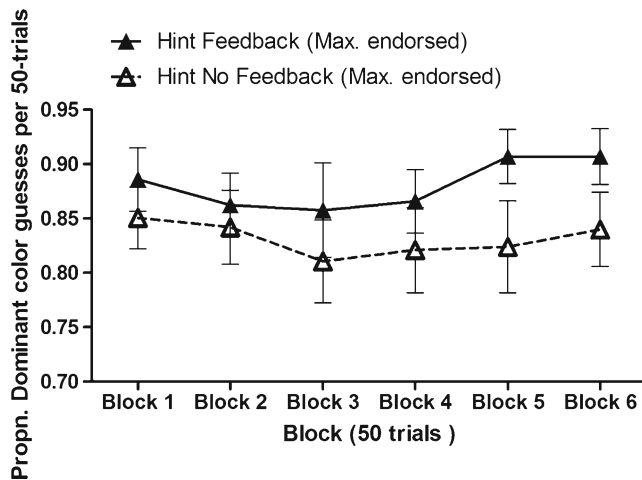
they endorsed, more participants prompted with the hint identified maximizing as superior (black bars in the upper panels) than spontaneously maximized when no hint was provided (far right open bars in lower panels). This effect is especially prominent in the no-hint no-feedback group, in which matching was overwhelmingly dominant in the first ten-roll game (lower right panel).

To investigate in more detail the interplay of feedback and hint over the entire 300 trials, we examined the prediction data across blocks for the subset of participants in the hint conditions who explicitly identified maximizing as optimal from the outset. Figure 5 plots these data and clearly indicates that there was still a small benefit of receiving outcome feedback across trials. In a Block  $\times$  Feedback ANOVA, this effect manifested itself as a significant linear interaction,  $F(1, 36) = 6.02$ ,  $p = .019$ ,  $\eta_p^2 = .143$ , indicating that the difference in maximizing as a function of feedback condition increased across blocks. No other effects in this analysis were significant. (The proportions of dominant-color predictions across blocks for participants endorsing matching were .77 and .71 for the hint feedback and hint no-feedback conditions, respectively.)

*Modeling the data* In a further effort to examine the effect of a hint on performance and learning, we applied a reinforcement-learning model to all of the data from the hint feedback and no-hint feedback conditions. We focused on

**Fig. 4** Numbers of dominant-color guesses made by participants in the first ten-roll game of Experiment 2. “Endorse Max” and “Endorse Match” refer to the strategies selected when participants were prompted with the hint (see text). (Note that “5” on the  $x$ -axis indicates five or fewer “green” guesses.)





**Fig. 5** Mean proportions of dominant-color guesses for each of the six blocks of 50 rolls in Experiment 2 for participants in the two hint conditions who endorsed maximizing as the optimal strategy when prompted with the hint before the prediction task. Error bars indicate SEMs

these two conditions because, as can be seen in Fig. 3, the conditions without feedback showed very little evidence of learning across blocks. Our key interest for this modeling exercise was to investigate how learning from feedback was affected by the presence/absence of a hint. Specifically, we wanted to see whether modeling could provide convergent evidence for the “top-down”/“bottom-up” interplay of the hint and feedback, respectively. To achieve this goal, we applied two advice-reinforcement combination (ARC) models developed by Biele et al. (2009), which we will discuss in the next section<sup>4</sup> (see also Yechiam & Busemeyer, 2005).

**The models** We considered two models: the ARC-initial and ARC-outcome models. These models are alike, in that they pertain to a decision environment in which participants could rely on both advice and individual learning experience when choosing their decision strategy. The models differ in terms of the temporal effect of advice on learning: ARC-initial assumes that the effect of advice is strongest at the earliest stage of learning and decays with time, whereas ARC-outcome assumes that the effect of advice on learning grows over time, so that it is more pronounced later on (Biele et al., 2009). The behavioral results from Experiment 2 suggest that the ARC-initial model should provide a better fit to the data because the hint (advice) appears to confer an early advantage (see Fig. 3).

Both models share the assumption that the participant enters the decision environment with an initial propensity for each of the two response options. Upon choosing a response option, the accuracy of the choice is used to update the propensity of the chosen option. Independent of the choice made, differential propensities to choose one response over

the other decay with time. Response sensitivity (updating in response to feedback) and decay rates are free parameters in both versions of the model.

ARC-initial allows for an initial propensity favoring the optimal response option. An initial bias  $\iota$  of 0 would indicate no bias, and positive values of initial bias  $\iota$  would indicate an a priori preference for the optimal response. In total, ARC-initial has three free parameters to estimate: response sensitivity  $\lambda$ , memory decay  $\varphi$ , and initial bias  $\iota$ . ARC-outcome does not allow for an initial propensity favoring the optimal response option. Instead, the reward rate for the optimal response, when it is correct, is enhanced by a free parameter  $\rho$  specifying additional reinforcement for choosing the recommended option. ARC-outcome also has three free parameters: response sensitivity  $\lambda$ , memory decay  $\varphi$ , and additional reinforcement  $\rho$ . A more detailed description of both models may be found in the Appendix.

**Model results** Individual parameter estimates were obtained for both the ARC-initial and ARC-outcome models for each of 52 participants (26 per condition). Full details of the model fitting may also be found in the Appendix. Most importantly, only the ARC-initial model successfully converged (i.e., obtained stable parameter estimates), so we will limit our discussion to the results of that model. The fact that only the ARC-initial model converged successfully lends further credibility to our claim that the effects of the hint are most pronounced at an early stage of learning.

The ARC-initial model had a higher parameter estimate of response sensitivity  $\lambda$  for the hint feedback condition (mean = 1.14, 95 % CI = 0.63–1.65) than for the no-hint condition (mean = 0.73, 95 % CI = 0.43–1.03) (Cohen’s  $d = 0.41$ ). This signifies that participants in the hint condition were more sensitive to differences in choice propensities. Recall that a response sensitivity  $\lambda$  of 0 corresponds to random guessing, whereas a response sensitivity  $\lambda$  approaching infinity corresponds to strictly consistent behavior. Participants in the hint condition, then, responded in a slightly more consistent—and hence, rational—fashion, given that the higher value indicates more maximizing. The parameter estimates of memory decay  $\varphi$  were similar between the two conditions (mean = 0.21, 95 % CI = 0.15–0.27, and mean = 0.17, 95 % CI = 0.14–0.20, for the hint feedback and no-hint feedback conditions, respectively), suggesting equal amounts of decay (Cohen’s  $d = 0.32$ ). This makes sense, because feedback was provided in the same fashion in both conditions. The third parameter,  $\iota$ , quantifies the initial propensity toward the optimal response, relative to the nonoptimal response. In our task, in which participants had all of the information necessary for identifying the optimal response at the outset—regardless of the hint—one would expect  $\iota > 0$  in both conditions. This was indeed the case, but in addition the parameter was estimated to be higher in the hint feedback condition (mean = 26.6, 95 % CI = 18.9–34.3) than in

<sup>4</sup> We thank Lael Schooler for suggesting the application of this model.

the no-hint feedback condition (mean = 18.2, 95 % CI = 10.5–26.0) (Cohen's  $d = 0.45$ ) This difference demonstrates that, on average, the hint increased the initial propensity toward maximizing; the fact that the difference was not large simply reflects that, regardless of the hint, some participants respond optimally and some do not (cf. the frequency distributions showing individual differences in optimal responding in Fig. 4).

The primary conclusion from this modeling exercise was that a model that characterizes “top-down” advice (i.e., the hint) as being combined early in learning with “bottom-up” reinforcement from outcome feedback provides a good account of the data. Moreover, the parameter estimates for this model differ in sensible, interpretable ways across the conditions. The fit of this model and the differences in parameter values across conditions are somewhat remarkable, given that the model was developed by Biele et al. (2009) for a task in which participants had to learn the contingencies associated with different response options—unlike our task, in which all of this information was readily available. It is, therefore, perhaps unsurprising that while the differences in parameter values were consistent with our account (e.g., higher  $\lambda$  and  $\iota$  in the hint feedback condition), the differences were not significant across conditions (all  $t$ s < 1.6).

## General discussion

This research sought to shed light on a puzzling finding: Despite having all of the information necessary from the outset to identify (and employ) maximizing as a superior strategy in a simple sequential-choice task, participants receiving outcome feedback steadily increase their maximizing rate across trials (Newell & Rakow, 2007). In Experiment 1, we found that the provision of a hint about the earning potentials of two different strategies had similar facilitative effects on maximizing rates (cf. Koehler & James, 2010), but we failed to find any evidence that outcome feedback increased maximization rates. This is inconsistent with our hypothesis that feedback helps some participants to “discover” the optimal strategy. The results of Experiment 2, however, suggested a somewhat more complicated interplay of feedback and strategy availability: Although the hint served to make maximizing more available in early trials, as the opportunities for learning increased (i.e., the number of trial-by-trial predictions surpassed that of Exp.1), outcome feedback *did* gradually begin to exert its influence on responding, even when a hint had been provided. This influence extended to those participants who had correctly identified maximizing as optimal when prompted with the initial hint (see Fig. 5).

The results clarify the roles of two sources of information that can help guide participants toward optimal responding in the die task. One source—the hint—appears to act on the initial expectancies or strategies that participants generate

(“top-down”) when they are faced with the description of the task (James & Koehler, 2011). Without being explicitly encouraged to consider ways of responding, a proportion of participants simply fail to realize that maximizing is the optimal response, and so are seduced into responding in a way that is “representative” of the probabilities underlying the task (cf. Kahneman & Tversky, 1972; Koehler & James, 2010). When a hint is provided, it pushes some of these participants toward maximizing from the outset (see Figs. 2 and 4).

The other source of information—outcome feedback—acts more gradually (“bottom-up”) and triggers a trial-by-trial search for alternative choice strategies that helps at least some participants discover the maximizing strategy. Importantly, this period of discovery takes time; the effect of outcome feedback was not apparent within the first 50 trials of either experiment. The application of a reinforcement learning model to our data supports these general conclusions. The better-fitting model was one that combined an initial propensity to choose the maximizing response outcome, which was stronger in the hint than in the no-hint condition, with a mechanism that learned gradually from feedback.

Another consistent pattern from these experiments is that when neither a hint nor feedback is provided, in the aggregate, many participants are prone to start with, and stay with, the matching strategy. One possible interpretation of this result (suggested by a reviewer) is that participants, in this condition in particular, misapprehend the requirements of the task (see, e.g., Hilton, 1995). Specifically, they might believe that they are being asked to produce a potential *outcome distribution* for the die rolls, rather than to make independent predictions for each trial. If this were the case, probability matching would be an appropriate strategy.

While this is possible, we think that this interpretation is unlikely for several reasons. First, the incentive structure of the task (payment for *each* correct prediction) clearly emphasizes the importance of independent predictions—and participants were reminded of this structure after every ten trials in Experiment 2. Second, a classic study by Goodnow (1955) demonstrated that although instructions to consider a two-choice task similar to ours as a problem-solving task (“find the pattern”) resulted in less maximizing than did an instruction to treat it as a “gambling task,” those given the latter frame still fell well short of maximizing. Third, when the die task is presented purely as a described problem (in which there are no prediction trials, and thus no ambiguity about the task requirements), a significant proportion of people still endorse probability matching as the optimal strategy (Newell & Rakow, 2007; West & Stanovich, 2003). Thus, the observed irrationality cannot purely reside in a failure to understand the task requirements.

Over the years, many other reasons have been proposed for the persistence of probability matching despite its sub-optimality, some of which may be applicable to our data.



Perhaps one of the most enduring is that the utility of predicting the rare event might outweigh the often small monetary benefit of maximizing, and might alleviate the boredom of continually making the same response (e.g., Goodnow, 1955; Shanks et al., 2002). In the words of one participant from the Goodnow study (in which the reinforcement was 70 % left, 30 % right): “It’s easy to win on the left; the real skill comes in winning on the right.” As was pointed out by Goodnow, “One could win on the right key by betting on it very often but... this would be no more a test of skill than deer-hunting with a machine gun” (p.113). While it is possible that some of our participants were attempting some “skilled” predictions of the rare event, the use of a familiar chance device (a die) with precise outcome probabilities (and no memory) ought to at least somewhat discourage pattern-seeking and other attempts at “skilled” prediction. Of course, many manipulations discourage “skill development” (substantial monetary rewards, extended practice, etc.; see Shanks et al. 2002); our contribution is to show that two such factors—a hint and feedback—operate via somewhat different mechanisms.

Naturally, though, even without the prompts provided by the hint and feedback, some participants do identify maximizing as the superior strategy. In Experiment 1, four participants (out of 31, 13 %) in the no-hint no-feedback condition predicted the dominant color on all 50 trials, and in Experiment 2, two participants (out of 23, 9 %) from this condition predicted the dominant color on all 300 trials (with a further two only making a suboptimal prediction once across all trials).

The presence of such individuals suggests that some fundamental cognitive abilities (over and above the impact of the independent variables) contribute to maximizing behavior in binary prediction. Some support for this interpretation comes from the correlations between participants’ scores on the Cognitive Reflection Test (CRT, a measure of controlled/deliberative thinking; Frederick, 2005) and their numbers of maximizing responses in the prediction task. On completion of the prediction task, participants in both experiments reported here answered the three-item CRT; the correlations between CRT scores and the numbers of dominant-color predictions were  $r = .33$  and  $.23$ ,  $ps < .05$ , in Experiments 1 and 2, respectively. This pattern of correlations is consistent with recent research showing that the CRT is a potent predictor of performance on a range of “heuristics-and-biases”-type tasks, including a fully described version of the die problem used here (Koehler & James, 2010; Toplak, West, & Stanovich, 2011). Taken together, these findings suggest that, for some individuals, the fully described die problem indeed yields maximizing, as would be expected from a rational agent. For others, though, the maximizing strategy does not come readily to mind in response to a description of the problem, but instead

must be “discovered” through the generation and trial-by-trial testing of alternative choice strategies.

### Appendix: Details of the model and model implementation

The ARC-initial and ARC-outcome models share the assumption that the decision maker (DM) enters the decision environment with an initial propensity for each of the two response options,  $q_1(t)$  and  $q_2(t)$ . Upon choosing a response option, the accuracy (and, as such, the reward) of the choice is used to update the propensity of the chosen option. Independent of choice,  $q_1(t)$  and  $q_2(t)$  decay with time. The probability to choose either of the two response options is defined as the odds of a function of  $q_1(t)$  and  $q_2(t)$  (see Eq. 2). After making a response,  $q_1(t)$  and  $q_2(t)$  are updated according to

$$\begin{aligned} q_1(t+1) &= (1 - \varphi) * q_1(t) + r(t), \\ q_2(t+1) &= (1 - \varphi) * q_2(t) + r(t), \end{aligned} \quad (1)$$

where  $\varphi$  is a free decay parameter determining the DMs memory for past experiences. A memory decay  $\varphi$  of 0 corresponds to perfect memory, whereas a  $\varphi$  of 1 corresponds to acting strictly on the outcome of the most recent trial. The quantity  $r(t)$  is the reward received if the option was chosen, and zero if the option was not chosen. The choice propensities determine the probability of choosing the optimal response (i.e., the dominant color) according to

$$p(t) = \exp[\lambda * q_1(t)] / \{\exp[\lambda * q_1(t)] + \exp[\lambda * q_2(t)]\}, \quad (2)$$

where  $\lambda$  is a free parameter specifying the DMs sensitivity to differences in choice propensities. A response sensitivity  $\lambda$  of 0 corresponds to random guessing, whereas a response sensitivity  $\lambda$  approaching infinity corresponds to strictly rational behavior (i.e., either maximizing or minimizing, depending on whether the DM has the correct choice propensities).

We implemented both the ARC-initial and ARC-outcome models in the Bayesian software program WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000).<sup>5</sup> For the parameters to be estimated, we used the following priors:  $\lambda \sim U(0, 5)$  and  $\varphi \sim U(0.1, 0.9)$ , where U indicates a uniform distribution, with the lower and upper bounds of the parameter in parentheses. Additionally,  $\iota \sim U(0, 100)$  for the ARC-initial model and  $\rho \sim U(0, 10)$  for the ARC-outcome model. For each parameter, we ran three separate chains, each of which consisted of 2,000 iterations, of which the first 1,500 were

<sup>5</sup> The model files can be accessed from D.v.R.’s website: [www.donvanravenzwaaij.com](http://www.donvanravenzwaaij.com).

treated as burn-in samples. Burn-in samples are meant to calibrate the sampling process and are not included in the posterior distribution (see Lee & Wagenmakers, 2011, for a more elaborate explanation of Bayesian modeling).

We examined whether the models converged successfully (i.e., whether stable parameter estimates could be obtained) by calculating an “*R*-hat” for each parameter (Gelman & Rubin, 1992). *R*-hat is a statistic that compares the sample variances between separate chains to the sample variances within the chains. When the chains are indistinguishable, so are the between- and within-sample variances, and *R*-hat equals 1. A guiding principle is that an *R*-hat higher than 1.05 is considered inadequate (no stable parameter estimates could be obtained). For the ARC-initial model, seven of the 156 parameters had an *R*-hat higher than 1.05, suggesting that the model converged successfully. In contrast, 80 out of the 156 parameters of the ARC-outcome model had an *R*-hat higher than 1.05, indicating that the ARC-outcome model did not converge successfully for these data.<sup>6</sup>

The next step in model evaluation was to assess model fit. As a means to assess model fit, we generated model predictives of the data based on the posterior distribution and compared these to the actual data. Using these so-called *posterior predictives*, we calculated the proportions of times the dominant color was predicted to be chosen for each participant. We then calculated the mean absolute deviations between the predictives and the data and compared these for the two models and for two simple heuristics: maximizing (picking the dominant color exclusively) and matching (picking the dominant color 70 % of the time). Doing so led to mean proportional discrepancies in dominant-color responses of 2.2 percentage points for the ARC-initial model, 0.3 percentage points for the ARC-outcome model, 14.5 percentage points for the maximizing model, and 15.8 percentage points for the matching model.

On the basis of the combination of model convergence and model fit, our preferred model of choice was the ARC-initial model, which both converged successfully and fit the data very well, as compared to the simple heuristics.

**Author note** The support of the Australian Research Council (Grant Nos. DP 110100797 and FT110100151) to the first author is gratefully acknowledged, as is funding from the Natural Sciences and Engineering Research Council of Canada to the second author. We thank Helen Archibald for assistance with data collection, and Christin Schulze and Lael Schooler for insightful comments on an earlier draft.

<sup>6</sup> Additionally, we implemented a hierarchical version of the ARC-initial model in which parameters were drawn from a group distribution per condition. This model did not converge successfully either, with 54 of the 162 parameters having an *R*-hat higher than 1.05.

## References

- Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science*, *33*, 206–242.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 24–42.
- Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*, 416–422.
- Gal, I., & Baron, J. (1996). Understanding repeated choices. *Thinking and Reasoning*, *2*, 81–98.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.
- Goodnow, J. J. (1955). Determinants of choice-distribution in two-choice situations. *The American Journal of Psychology*, *68*, 106–116.
- Hermstein, R. J. (1997). *The matching law: Papers in psychology and economics* (H. Rachlin & D. I. Laibson, Eds.). Cambridge, MA: Harvard University Press.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *The Behavioral and Brain Sciences*, *24*, 383–403.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248–271.
- James, G., & Koehler, D. J. (2011). Banking on a bad bet: Probability matching in risky choice is linked to expectation generation. *Psychological Science*, *22*, 707–711.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.
- Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113*, B123–B127.
- Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, *38*, 667–676.
- Lee, M. D., & Wagenmakers, E.-J. (2011). A course in Bayesian graphical modeling for cognitive science. Available at <http://www.ejwagenmakers.com/BayesCourse/BayesBook.pdf>
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.
- Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). *Straight choices: The psychology of decision making*. Hove: Psychology Press.
- Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review*, *14*, 1133–1139.
- Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*, 274–279.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics and biases tasks. *Memory & Cognition*, *39*, 1275–1289.
- Vulkan, N. (2000). An economist’s perspective on probability matching. *Journal of Economic Surveys*, *14*, 101–118.
- West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, *31*, 243–251.
- Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*, 387–402. doi:10.3758/BF03193783