

EMPIRICAL ARTICLE

Probability matching and statistical naïveté

Megan Barlow, Tiffany Doan, Ori Friedman  and Stephanie Denison

Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Corresponding authors: Stephanie Denison; Email: stephanie.denison@uwaterloo.ca or Ori Friedman;
Email: friedman@uwaterloo.ca

Received: 4 July 2025; **Revised:** 4 September 2025; **Accepted:** 6 September 2025

Keywords: probabilistic reasoning; statistical inference; probability matching; choice task

Abstract

People often probability match: they select choices based on the probability of outcomes. For example, when predicting 10 individual results of a spinner with 7 green and 3 purple sections, many people choose green mostly but not always, even though they would be better off always choosing it (i.e., maximizing). This behavior has perplexed cognitive scientists for decades. Why do people make such an obvious error? Here, we provide evidence that this difficulty may often arise from statistical naïveté: Even when shown the optimal strategy of maximizing, many people fail to recognize that it will produce better payouts than other strategies. In 3 preregistered experiments ($N = 907$ Americans tested online), participants made 10 choices in a spinner game and estimated the payout for each of 3 strategies: probability matching, maximizing, and 50/50 guessing. The key finding across experiments is that while most maximizers recognize that maximizing results in higher payouts than matching, probability matchers predict similar payouts for each.

1. Introduction

Suppose you notice that each weekend, an ice cream truck visits a large park in your town 70% of the time and a smaller park 30% of the time. If you hope to encounter the truck, would you go to the larger park 70% of the time and the smaller 30% of the time? Hopefully, not. You should go to the larger park each weekend. After all, you cannot predict when it will go to each park with certainty. Yet, people often probability match. They predict the events in approximate proportion to the probability of their occurrence (e.g., Estes and Straughan, 1954; Goodnow, 1955; Koehler and James, 2009, 2010, 2014; Neimark and Shuford, 1959; Vulkan, 2000) even though this results in lower predictive accuracy than an alternative strategy of maximizing—predicting the most probable event every time.¹

On one account, probability matching arises because people search for patterns in data (e.g., Gaissmaier and Schooler, 2008). On another account, probability matching arises because people are tempted by an intuitive, but unwise strategy of reproducing the expected frequencies in their choices, and they do not consider using other, better strategies like maximizing (e.g., Koehler and James, 2009). Here, we propose a further explanation, which is not mutually exclusive to existing accounts. Specifically, people may probability match because they are naïve to the expected payout of potential strategies. This proposal is broadly consistent with theoretical accounts of decision making

¹Recent work has questioned how often probability matching happens in probability *learning* tasks, where participants learn probabilities over numerous trials (Montag, 2021; Saldana et al., 2022). Although we review work from this literature, our focus is on tasks that do not require learning—tasks where participants know the probabilities upfront.

that suggest many people do not have the computational abilities (either due to poor computational skills or cognitive miserliness) to come to optimal solutions (see Kahneman, 2011).

On the pattern-search account, probability matching arises from people overapplying a strategy that is sometimes useful, specifically when outcomes are patterned. For example, if you noticed that the ice cream truck visited the smaller park every third week, you should no longer go to the larger park each week. Hence, probability matching may arise because people are inclined to search for patterns (Gaissmaier and Schooler, 2008; Wolford et al., 2004). Indeed, participants in an experimental task who probability matched when they should not (i.e., when they were explicitly told that outcomes were produced randomly), were more likely to find patterns in other situations where they did exist (Gaissmaier and Schooler, 2008). While these findings suggest that probability matching may have a ‘smart’ utility, the pattern search account cannot tell the whole story. This is because probability matching occurs at *similar rates* whether patterns can be conceivably detected or not (Koehler and James, 2009).

Why do people probability match even when patterns cannot be detected? One view is that matching is a more readily available strategy than maximizing (Koehler and James, 2009, 2010; Kogler and Kühberger, 2007; West and Stanovich, 2003). This account is often associated with a dual-system framework, where probability matching reflects the fast, intuitive responses of System 1 and maximizing reflects the slower, more effortful responses of System 2. Probability matching likely comes to mind as a System 1 response due to attribute substitution (Kahneman and Frederick, 2002), where people answer the difficult question ‘How many times should X and Y be predicted?’, by answering the easier question ‘How many times are X and Y expected?’ (Koehler and James, 2009, 2010).

Consistent with this, people often probability match when they play a game where a spinner with 7 green and 3 purple sections is spun 10 times—that is, a game where it is easy to generate aggregated predictions about the number of times each color is expected. But they match less when they play 10 different games (e.g., 10-sided die, 10 section spinner)—a situation where aggregating is less natural (James and Koehler, 2011). Additionally, when participants must explicitly evaluate probability matching versus maximizing as strategies, or when the experimenter brings both strategies to participants’ attention, people endorse and engage in more maximizing (Koehler and James, 2009, 2010; for other kinds of evidence favoring this dual systems account of probability matching, see Fantino and Esfandiari, 2002; Kogler and Kühberger, 2007; Newell et al., 2013; Taylor et al., 2012).

We examine a further explanation for why people probability match. People may often do this because of statistical naïveté—they may fail to estimate the payout for strategies like probability matching and maximizing. For example, consider a task where you guess the color on each of 10 spins from a spinner with 7 green sections and 3 purple ones. Many people may struggle to see that choosing green all 10 times will lead (on average) to 7 correct guesses. They might also struggle to recognize that guessing green fewer times (e.g., 7 of 10) will typically lead to poor performance.

Some previous findings are broadly consistent with this proposal. A preference for maximizing over matching (and over other suboptimal strategies) is predicted by greater numeracy (Corser et al., 2024), and perhaps by having taken more math and statistics courses (Gal and Baron, 1996; but for conflicting findings, see Rakow et al., 2010; West and Stanovich, 2003). Also, in some experiments, participants were told about both strategies and then asked which would yield better results (Gal and Baron, 1996; Koehler and James, 2010; Newell et al., 2013). Although most participants indicated that maximizing was better, many did not. For instance, in Newell et al., 64% of participants in one experiment and 74% in another recognized that maximizing was better. However, even these figures might provide an inflated sense of people’s recognition that maximizing is better. If participants had not been asked about which strategy is better, many might have assumed that both would produce similar results.²

²Statistical naïveté might also contribute to the choices of people who seem to grasp the benefits of maximizing. People are more likely to maximize if they are presented with both strategies before participating in the task. But some people might adopt the strategy without understanding or believing that it is actually better—some might try it simply because it was suggested.

A clearer picture of participants' expectations might be provided, though, by having participants predict the payout of each strategy. We did this in 3 experiments. In each, we asked participants to imagine that a character, Pat, was playing a game with a spinner segmented into 7 green and 3 purple sections. There would be 10 spins and Pat would win a quarter for each correct guess. We asked participants about where Pat should look for the quarters to gauge whether participants themselves probability matched, maximized, or heeded some other strategy. We also told them to imagine that hundreds of people had played the game and produced the following strategies: matching (7 green, 3 purple); maximizing (10 green); and 50/50 responding (5 green, 5 purple) and to state how many quarters people using those strategies would win on average.

We expected that participants who themselves maximized would generally predict fewer wins for the matching than maximizing strategy. These participants might recognize, for instance, that with 7:3 odds, maximizing should typically result in 7 wins, whereas matching must lead to fewer wins on average (i.e., since guesses of the minority color on average will mostly fall on the wrong trials).³ The statistical naïveté account, though, predicts that matchers will not show this pattern and might instead offer similar rates of success for matching and maximizing alike.

2. General methods

All experiments were preregistered. Preregistrations, materials, data, and analysis code are at <https://osf.io/9kbnx/>. Participants in all experiments were recruited from Cloud Research (Litman et al., 2017). They were located throughout the United States and had a HIT rate above 95%. In all experiments, we sought to test at least 100 participants per between-subjects conditions. We chose this number because we have found it adequate in previous work using related designs. This research submitted under the name, 'Inferences about different kinds of outcomes' (ORE#31953), received ethics clearance through a University of Waterloo's Research Ethics Board.

Participants completed the experiments online using Qualtrics. In each experiment, participants first read some preliminary information about a guessing game, and were then asked two 4-option multiple-choice comprehension questions about it. If they answered either question incorrectly, the instructions repeated and the comprehension question repeated. If participants failed this 3 times, the survey continued, but data were excluded. After completing the main task in each experiment, participants were asked another attention check question. Participants were also excluded if they answered it incorrectly.

3. Experiment 1

3.1. Participants

We tested 226 participants ($M_{\text{age}} = 39$, 86 female, 139 male, 1 other/prefer not to say). Six additional participants were excluded based on the preregistered exclusion criteria (i.e., failing 3 attempts of the pretest comprehension checks or failing the posttest comprehension check).

3.2. Materials and procedures

First, participants read a description of a game in which a contestant, Pat, will look for quarters under pairs of green and purple cups, with a goal of finding as many quarters as possible (see Figure 1). A quarter was hidden under 1 cup in each pair (10 quarters total), with the hiding location determined by 10 spins of a spinner. The spinner had 7 green sections and 3 purple sections.

On a next screen, participants were asked to indicate where Pat should look by selecting 'Green' or 'Purple' for each of the 10 pairs (see Figure 2).

³We did not expect that maximizers would calculate the precise number of wins that matching would produce on average (i.e., 5.8). Indeed, our method was not well designed for testing this, since we forced participants to offer whole numbers as answers.

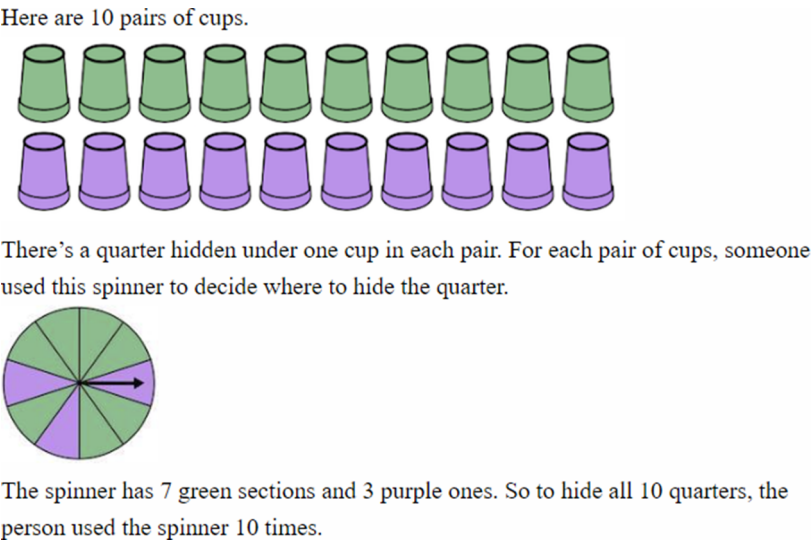


Figure 1. Instructions to the game for all experiments.

For each pair, please indicate where Pat should look.








	GREEN	PURPLE
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>

Figure 2. Choice screen for Pat in Experiment 1.

Participants then read that hundreds of people did this activity, and that on the next screens they would see what some people did. Across the next 3 screens, participants saw the following strategies one at a time: probability matching, maximizing, and 50/50. They were asked to indicate, on average, how many quarters people would find using that strategy on an 11-point scale, ranging from 0 to 10 (see Figure 3).

Following this, participants were asked a final 4-option multiple-choice question about what was hidden under the cups.

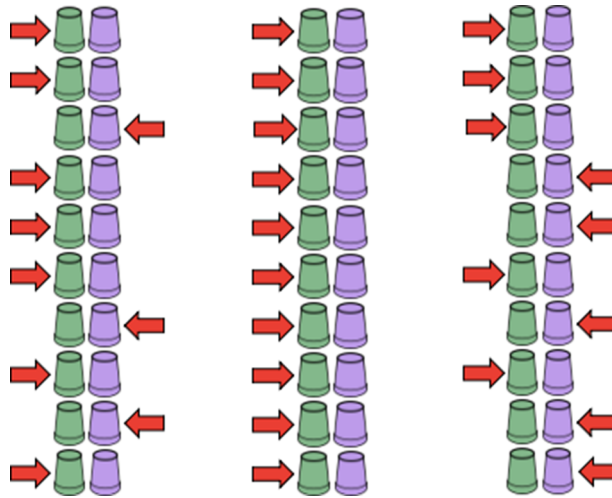


Figure 3. The depiction of the 3 strategies for all experiments.

Note: Each strategy was shown on a separate screen with a scale below for participants to indicate the number of quarters the strategy would yield (0–10).

3.3. Results

Based on responses about what Pat should choose, participants were categorized as maximizers (9 or 10 green), matchers (6, 7, or 8 green),⁴ or other (5 or fewer green). Overall, 43% of participants were matchers, 45% were maximizers, and 12% showed some other pattern.

We examined whether those who probability match or maximize differ in their expectations of the outcomes (i.e., the number of quarters won) for various strategies; see Figure 4. A 2 (Category: Matcher, Maximizer) \times 3 (Strategy: Matching, Maximizing, 50/50) analysis of variance (ANOVA) revealed a significant interaction, $F(1.86, 366.97) = 29.49, p < .001$. Post hoc comparisons revealed that matchers expected similar outcomes for both matching ($M = 5.87, 95\% \text{ CI } [5.55, 6.19]$) and maximizing ($M = 5.98, 95\% \text{ CI } [5.64, 6.32]$) strategies, $t(197) = 0.69, p > .770$, but recognized that both would result in better outcomes than 50/50 ($M = 4.46, 95\% \text{ CI } [4.17, 4.75]$), $t(197) = 9.24, p < .001$, $t(197) = 8.00, p < .001$, respectively. Maximizers correctly indicated that maximizing ($M = 6.86, 95\% \text{ CI } [6.53, 7.19]$) would yield a better outcome than matching ($M = 5.18, 95\% \text{ CI } [4.86, 5.49]$), $t(197) = 10.50, p < .001$, and that both would result in better outcomes than 50/50 ($M = 3.75, 95\% \text{ CI } [3.47, 4.03]$), $t(197) = 9.50, p < .001$ and $t(197) = 16.60, p < .001$, respectively.

These results tell us about the aggregate predictions of matchers and maximizers. In an additional nonpreregistered analysis, we also looked at their individual responses. We categorized participants as giving correct predictions if their predicted payouts were ordered maximizing > matching > 50/50; otherwise, they were categorized as incorrect. Correct predictions were more often given by maximizers (77/101 = 76%) than matchers (30/98 = 31%), $\chi^2(1) = 41.65, p < .001$.

3.4. Discussion

Maximizers recognize that maximizing is better than matching and matching is better than 50/50 guessing when predicting the average payout of hundreds of players. Conversely, matchers appear to think that matching and maximizing are similarly good strategies, suggesting they have difficulty determining the value of maximizing.

⁴As preregistered, this is how probability matchers and maximizers have been categorized in prior work (e.g., Koehler and James, 2010).

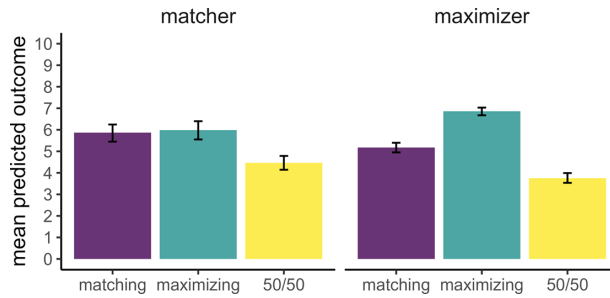


Figure 4. Mean predicted outcomes in Experiment 1.

Note: Participants categorized as matchers (left) and maximizers (right) predicted the outcomes (0–10) for 3 strategies (matching, maximizing, 50/50). In all graphs, error bars show 95% CI.

However, a nontrivial number of maximizers did not give correct predictions. Also, although it was not a focus of our analyses, maximizers (taken as a group) appeared to underestimate the payouts for both matching and 50/50 responding—with both predictions, the 95% confidence intervals did not include the correct expected rates of success (i.e., 5.8 for matching and 5.0 for 50/50). The response format of our experiment may have inadvertently caused this. To indicate what Pat should do, participants had to separately respond for each pair of cups. The least effortful strategy would be to click straight down one side, allowing some unmotivated participants to be classified as maximizers. In Experiment 2, we sought to replicate the findings but using a different response format.

4. Experiment 2

4.1. Participants

We tested 228 participants ($M_{\text{age}} = 41$, 93 female, 132 male, 3 other/prefer not to say). Two additional participants were excluded for failing the posttest comprehension check.

4.2. Materials and procedure

Experiment 2 was identical to Experiment 1 except participants were asked to indicate which color cup Pat should choose in each pair by inputting numbers into 2 text fields using a constant sum format. That is, in one field they indicated the number of green cups Pat should choose and in the other, the number of purple cups. The order of the text fields was randomized across participants.

4.3. Results

Again, participants were categorized based on their choices for Pat. Overall, 62% of participants were matchers, 19% were maximizers, and 19% showed some other pattern.

We examined whether matchers and maximizers differed when predicting payouts for each strategy; see Figure 5. A 2 (Category: Matcher, Maximizer) \times 3 (Strategy: Matching, Maximizing, 50/50) ANOVA revealed a significant interaction, $F(1.95, 356.62) = 9.33, p < .001$. Post hoc comparisons again revealed that matchers expected similar outcomes for both matching ($M = 5.77$, 95% CI [5.50, 6.05]) and maximizing ($M = 6.04$, 95% CI [5.78, 6.29]) strategies, $t(183) = 1.80, p = .174$, but indicated that both would yield more quarters than 50/50 ($M = 4.35$, 95% CI [4.13, 4.58]), $t(183) = 10.61, p < .001$, $t(183) = 10.91, p < .001$, respectively. Maximizers correctly indicated that maximizing ($M = 6.70$, 95% CI [6.26, 7.15]) would lead to a better outcome than matching ($M = 5.39$, 95% CI [4.90, 5.87]), $t(183) = 5.04, p < .001$, and that both would yield more quarters than 50/50 ($M = 3.86$, 95% CI [3.46, 4.27]), $t(183) = 6.36, p < .001$, $t(183) = 10.30, p < .001$, respectively.

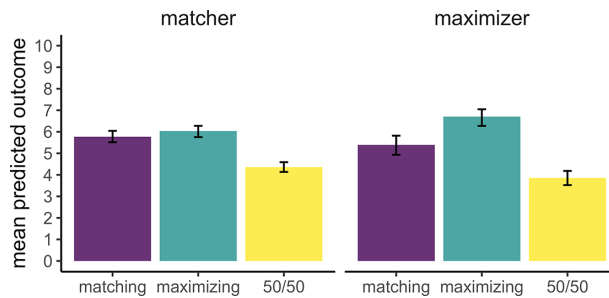


Figure 5. Mean predicted outcomes in Experiment 2.

Note: Participants categorized as matchers (left) and maximizers (right) predicted the outcomes (0–10) for 3 strategies (matching, maximizing, 50/50).

We again ran a nonpreregistered analysis in which participants were categorized as giving correct predictions if their predicted payouts were ordered maximizing > matching > 50/50. Correct predictions were more often given by maximizers (30/44 = 68%) than matchers (46/141 = 33%), $\chi^2(1) = 17.52, p < .001$.

4.4. Discussion

These findings replicate Experiment 1, suggesting that, in general, matchers see matching and maximizing as similarly good strategies, while maximizers recognize the superiority of maximizing over matching. As before, though, the predictions of some maximizers suggested they did not recognize the relative strengths of the strategies, and as a group they underestimated the odds for the 50/50 strategy (i.e., 95% confidence intervals for this strategy did not include 0.50).

In both experiments so far, participants made choices in the binary choice task (i.e., by saying what Pat should do) before predicting payouts of the different strategies. In Experiment 3, we examined whether participants are more likely to maximize if they see the strategies and predict payouts first.⁵ We also examined if some participants adopt maximizing after being introduced to it, but without recognizing that it will produce a better payout than matching.

5. Experiment 3

5.1. Participants

We tested 453 participants ($M_{\text{age}} = 42$, 223 female, 224 male, 6 other/prefer not to say). Nine additional participants were excluded for failing 3 attempts of the pretest comprehension checks or failing the posttest comprehension check.

5.2. Methods and procedure

Participants read the same game description as in Experiments 1 and 2 and were then randomly assigned to 1 of 2 conditions.

Those in the ‘before-and-after’ condition first completed all the steps from Experiment 2: they made choices for Pat, then saw the 3 strategies, and indicated the number of quarters people would get on average when using the strategies. After this, participants were presented with 10 pairs of cups again and were instructed to make choices for a new contestant, Jordan, exactly as they had done for Pat.

⁵This approach was adapted from Koehler and James (2010) and Newell et al. (2013). But while they asked participants a forced choice question about which strategy was better (i.e., an approach that implies that one is better), we separately gauged participants’ predictions about the outcomes of each approach.

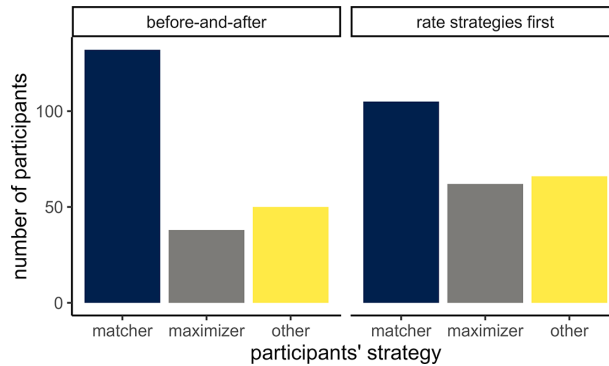


Figure 6. Number of participants showing each strategy across conditions in Experiment 3.

Note: Numbers of participants categorized as matchers, maximizers, or as showing some other strategy, on the basis of choices made either before they predicted payouts (before-and-after) or after this (rate strategies first).

In the ‘rate strategies first’ condition, participants completed the protocol from Experiment 2, but in reverse order: They first indicated the number of quarters each of the 3 strategies would yield, then made choices for Pat.

5.3. Results

Based on responses about *Pat*, participants were categorized as matchers (52%), maximizers (22%), or as using another strategy (26%). In line with the preregistrations, this experiment retained participants using other strategies in the main analyses (whereas these participants had not been examined in the previous experiments).

We first examined whether the proportion of participants employing each strategy differed across the between-subjects conditions. These analyses focus on predictions for *Pat*, as half of the participants were made aware of the strategies before making choices for *Pat* (strategies first condition) and half saw the strategies after choosing for *Pat* (before-and-after condition). This did indeed differ when looking at all 3 possible strategies (matching, maximizing, other), $\chi^2(2) = 10.68, p = .005$. As Figure 6 shows, matching was the most prevalent strategy in both conditions, but this tendency was more pronounced when choices were made for *Pat* before seeing the strategies than when strategies were presented first. Also, maximizing was relatively more prevalent when participants saw the strategies before choosing for *Pat* than when seeing the strategies after making choices for *Pat*.

We next focused on the before-and-after condition. In this condition, participants first indicated what *Pat* should do in the game, then predicted outcomes for each of the 3 strategies, and finally indicated what *Jordan* should do in the game. Examining responses from this condition allows us to look at participants who initially matched when choosing for *Pat*, but then switched to maximizing after predicting the payouts for the 3 strategies. These participants might switch because once they are introduced to maximizing, they recognize it as the best strategy. Alternatively, they might switch without understanding this.

To investigate these possibilities, we looked at participants who initially advocated matching, and then categorized them based on their predictions of the payouts for each strategy. They were categorized as giving correct predictions if their predicted payouts were ordered maximizing > matching > 50/50; otherwise, they were categorized as incorrect. Figure 7 (left panel) shows how many (initial) matchers in each category went on to say that *Jordan* should match, maximize, or do otherwise. A Fisher’s exact test revealed that matchers who were correct predictors were more likely than incorrect predictors to switch to maximizing, $p = .009$. However, as the figure also shows, even among the correct predictors, most persisted in matching!

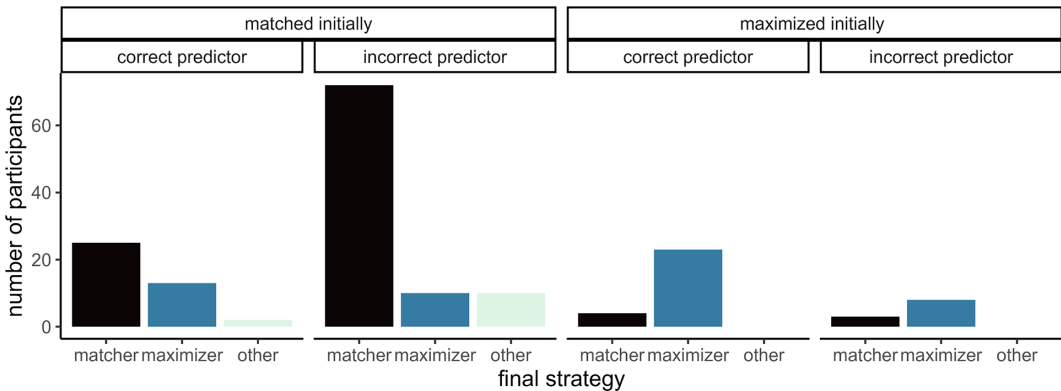


Figure 7. Number of participants showing each strategy after predicting payouts in Experiment 3.
 Note: Bars show how many participants in the before-and-after who initially matched (left) or maximized (right) went on to say another agent should match, maximize, or do otherwise.

In a nonpreregistered analysis, we also looked at participants who initially advocated maximizing. Figure 7 (right panel) shows the number of maximizers giving correct and incorrect predictions who advocated matching, maximizing, or using some other strategy when indicating what Jordan should do. Here, a Fisher's exact test did not reveal a significant effect, and most maximizers persisted in maximizing, $p = .474$.

Finally, in a further nonpreregistered analysis, we compared predicted payouts for each strategy across the 2 between-subjects conditions, which differed in whether participants completed the choice task before giving their predictions (before-and-after) or gave their predictions upfront (rate strategies first). An effect of condition could indicate that participants modified their predictions to fit their initial choices, for instance, increasing predictions of success for the strategy they used (or the one most closely resembling it). However, a 2 (Condition: Before-and-after, Rate strategies first) \times 3 (Strategy: Matching, Maximizing, 50/50) ANOVA revealed only a main effect of Strategy, $F(1.82, 820.33) = 174.59, p < .001$; the main effect of Condition was not significant, $F(1, 451) = 0.50, p = .479$, and the interaction was also not, $F(1.82, 820.33) = 2.45, p = .092$.

5.4. Discussion

In sum, although exposure to the strategies reduced the likelihood of probability matching, matching remained a common strategy. Also, we found that of matchers who switched strategies, many did not seem to understand the benefits of doing so.

6. General discussion

In 3 experiments, we found that participants who probability match are worse than maximizers in recognizing the relative strengths of different strategies. In Experiments 1 and 2, participants who probability matched generally thought that matching and maximizing would produce similar payouts. Participants who maximized, by contrast, did see maximizing as the superior strategy, though they too showed some signs of difficulty anticipating the payouts. In our third experiment, participants saw the strategies, and predicted the payouts, before participating in the choice task. This exposure improved performance: Participants were more likely to maximize if they completed the choice task after, rather than before exposure to the strategies. Nonetheless, matching remained the dominant strategy. Also, many participants who switched from matching to maximizing did not recognize that maximizing would produce better payouts.

Together these findings suggest that, for many people, probability matching arises in part from statistical naïveté, at least in tasks like ours where participants are informed of the probabilities upfront. This claim differs from what is suggested by the ‘availability’ explanation for probability matching (Koehler and James, 2009, 2010; Kogler and Kühberger, 2007; West and Stanovich, 2003). On that account, people match because this strategy occurs to them, whereas the maximizing strategy does not. Our claim is not about which strategies people are more or less likely to consider, but instead about whether people recognize the benefits of the strategies. And in this regard, our findings suggest that for many people, generating maximizing as a potential strategy is not much help. Even when it is pointed out to people (and they adopt it), they do not recognize its benefit. In line with this, some maximizers in our experiments did not understand its benefit—in every experiment, a nontrivial minority of maximizers gave incorrect predictions when predicting payouts.

It is an open question whether people, faced with the decision to match or maximize, *spontaneously* consider expected payouts or the odds of success (i.e., when not asked to do so). Some work suggests they might not. When participants have been asked to explain their choices, or to rate reasons for their choices, they rarely mentioned having calculated expected results, and this was even true for participants who maximized (Gal and Baron, 1996). Even so, it is difficult to see how people could grasp the superiority of maximizing over matching without *some* recognition that it would produce better outcomes.

A further open question is whether our findings are relevant to situations where people are not informed about the probabilities upfront, and instead learn them over time (e.g., Montag, 2021; Saldana et al., 2022). Differences might be expected given claims that probabilistic and statistical reasoning is more difficult when probabilities are described rather than experienced (i.e., the description–experience gap; e.g., Schulze and Hertwig, 2021). Nonetheless, the findings likely do extend to learning tasks (i.e., where probabilities are experienced rather than described): Maximizers in these tasks are more likely than others to endorse maximizing as the best strategy when asked about versions of the task where probabilities are described (Rakow et al., 2010).⁶

Although our work focuses on probability matching and binary choice tasks, the findings have broader import. People use statistical reasoning almost everywhere: from trying to choose the fastest line at the grocery store to deciding whether to implement costly national healthcare initiatives. Our tasks required relatively simple statistical reasoning. Even if it might be difficult for people to calculate the exact payouts offered by each strategy, the recognition that maximizing will produce better odds than matching is not complex. In our tasks, there were no complicated risk–reward structures, nor were there probabilities to update or potential patterns to detect—there was just a 70% chance of guessing correctly by choosing green for every spin. Nonetheless, many participants failed to recognize this. We have referred to this failure as stemming from statistical naïveté, or poor explicit statistical reasoning. This proposed mechanism may relate to numeracy, which impacts many judgment and decision-making tasks (e.g., ratio bias, denominator neglect, framing effects, gambler’s fallacy; e.g., Peters et al., 2006), including probability matching (Corser et al., 2024).

Our findings are also relevant to the classic debate about whether limitations and weaknesses in statistical reasoning are largely a byproduct of asking people to reason about percentages rather than natural frequencies (Cosmides and Tooby, 1996; Evans et al., 2000; Gigerenzer and Hoffrage, 1995; Kahneman and Tversky, 1996). This debate has typically concerned people’s ability to solve problems that require Bayesian reasoning. For example, problems asking people the likelihood of a medical test revealing a disease given the prior odds of the disease and the reliability of the test. However, we only showed participants natural frequencies: 7 green sections and 3 purple sections on a spinner—there was no mention of proportions or percentages whatsoever. Moreover, our tasks involved statistical problems

⁶The findings might not extend, though, to learning tasks where participants complete trials with feedback across many blocks, and with high incentives for correct responses. In such tasks, participants who initially match increasingly adopt maximizing as their strategy (e.g., Saldana et al., 2022; Shanks et al., 2002). In such tasks, participants likely come to understand the strategies differ in their payouts.

far tamer than those involving Bayesian reasoning. As such, our findings show that difficulties with statistical reasoning are widespread and are hardly eliminated by using natural frequencies.

Data availability statement. Preregistrations, materials, data, and analysis code are at <https://osf.io/9kbnx/>.

Funding statement. This work was supported by separate Discovery Grants from the Natural Sciences and Engineering Research Council of Canada awarded to SD and OF.

Competing interest. The authors declare none.

Ethical standards. Studies received approval from the Office of Research Ethics at the University of Waterloo.

Consent to participate. All participants consented to participate.

Consent for publication. N/A, the MS does not include individuating information about the participants.

References

- Corser, R., Voss, R. P., Jr., & Jasper, J. D. (2024). Do errors on classic decision biases happen fast or slow? Numeracy and decision time predict probability matching, sample size neglect, and ratio bias. *Journal of Numerical Cognition*, 10, e12473. <https://doi.org/10.5964/jnc.12473>.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1–73. [https://doi.org/10.1016/0010-0277\(95\)00664-8](https://doi.org/10.1016/0010-0277(95)00664-8).
- Estes, W. K., & Straughan, J. H. (1954). Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology*, 47(4), 225–234. <https://doi.org/10.1037/h0060989>.
- Evans, J., Handley, S. J., Perham, N., Over, D. E. & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, 77, 197–213. [https://doi.org/10.1016/S0010-0277\(00\)00098-6](https://doi.org/10.1016/S0010-0277(00)00098-6).
- Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology*, 56(1), 58–63. <https://doi.org/10.1037/h0087385>.
- Gaissmaier, W., & Schooler, J. L. (2008). The smart potential behind probability matching. *Cognition*, 109(3), 416–422. <https://doi.org/10.1016/j.cognition.2008.09.007>.
- Gal, I. & Baron, J. (1996). Understanding repeated simple choices. *Thinking and Reasoning*, 2, 81–98. <https://doi.org/10.1080/135467896394573>.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684–704. <https://doi.org/10.1037/0033-295X.102.4.684>.
- Goodnow, J. J. (1955). Determinants of choice-distribution in two-choice situations. *The American Journal of Psychology*, 68(1), 106–116. <https://doi.org/10.2307/1418393>.
- James, G., & Koehler, D. J. (2011). Banking on a bad bet: Probability matching in risky choice is linked to expectation generation. *Psychological Science*, 22(6), 707–711. <https://doi.org/10.1177/0956797611407933>.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3), 582–591. <https://doi.org/10.1037/0033-295X.103.3.582>.
- Kahneman, D. (2011). Thinking, fast and slow. *Farrar, Straus and Giroux*.
- Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition vs deliberation. *Cognition*, 113(1), 123–127. <https://doi.org/10.1016/j.cognition.2009.07.003>.
- Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, 38(6), 667–676. <https://doi.org/10.3758/MC.38.6.667>.
- Koehler, D. J., & James, G. (2014). Probability matching, fast and slow. In Ross, B. H. (Ed). *Psychology of Learning and Motivation*, 61 (pp. 103–131). Academic Press. <https://doi.org/10.1016/B978-0-12-800283-4.00003-4>.
- Kogler, C., & Kühberger, A. (2007). Dual process theories: A key for understanding the diversification bias?. *Journal of Risk and Uncertainty*, 34, 145–154. <https://doi.org/10.1007/s11166-007-9008-7>.
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behaviour Research Methods*, 49, 433–442. <https://doi.org/10.3758/s13428-016-0727-z>.
- Montag, J. L. (2021). Limited evidence for probability matching as a strategy in probability learning tasks. In K. D. Federmeier (Ed.), *The psychology of learning and motivation* (pp. 233–273). Elsevier Academic Press. <https://doi.org/10.1016/bs.plm.2021.02.005>.
- Neimark, E. D., & Shuford, E. H. (1959). Comparison of predictions and estimates in a probability learning situation. *Journal of Experimental Psychology*, 57(5), 294–298. <https://doi.org/10.1037/h0043064>.

- Newell, B. R., Koehler, D. J., James, G., Rakow, T., & van Ravenzwaajj, D. (2013). Probability matching in risky choice: The interplay of feedback and strategy availability. *Memory & Cognition*, 41, 329–338. <https://doi.org/10.3758/s13421-012-0268-3>.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2006.01720.x>.
- Rakow, T., Newell, B. R., & Zougkou, K. (2010). The role of working memory in information acquisition and decision making: Lessons from the binary prediction task. *Quarterly Journal of Experimental Psychology*, 63, 1335–1360. <https://doi.org/10.1080/17470210903357945>.
- Saldana, C., Claidière, N., Fagot, J., & Smith, K. (2022). Probability matching is not the default decision making strategy in human and non-human primates. *Scientific Reports*, 12(1), 13092. <https://doi.org/10.1038/s41598-022-16983-w>.
- Schulze, C., & Hertwig, R. (2021). A description–experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups. *Cognition*, 210, 104580. <https://doi.org/10.1016/j.cognition.2020.104580>.
- 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(3), 233–250.
- Taylor, E. G., Landy, D. H., & Ross, B. H. (2012). The effect of explanation in simple binary decision tasks. *Quarterly Journal of Experimental Psychology*, 65(7), 1361–1375. <https://doi.org/10.1080/17470218.2012.6566>.
- Vulkan, N. (2000). An economist’s perspective on probability matching. *Journal of Economic Surveys*, 14(1), 101–118. <https://doi.org/10.1111/1467-6419.00106>.
- West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, 31, 243–251. <https://doi.org/10.3758/BF03194383>.
- Wolford, G., Newman, S. E., Miller, M. B., & Wig, G. S. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(4), 221–228. <https://doi.org/10.1037/h0087446>.