Received: 18 April 2017 | Revised: 2 March 2018 | Accepted: 2 March 2018

DOI: 10.1002/bdm.2080

RESEARCH ARTICLE

WILEY

Switch rates do not influence weighting of rare events in decisions from experience, but optional stopping does

Kevin W. Soo | Benjamin M. Rottman

Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence
Kevin W. Soo, Department of Psychology, University of Pittsburgh, LRDC 720, 3939 O'Hara Street, Pittsburgh, PA 15260, USA.
Email: kevin.soo@pitt.edu

Funding information
National Science Foundation (NSF), Grant/Award Number: 1430439

Abstract

The current research investigates how people decide which of two options produces a better reward by repeatedly sampling from the options. In particular, it investigates the roles of two features of search, optional stopping and switch rate, on participants' final judgments of which option is better. First, in two studies, we found evidence for a new optional stopping effect; when participants stopped sampling right after experiencing a rare outcome, they made decisions as if they overweighted the rare outcome. Second, we investigated an effect proposed by Hills and Hertwig (2010) that people who frequently switch between options when sampling are more likely to make decisions consistent with underweighting rare outcomes. We conducted a theoretical analysis examining how switch rate can influence underweighting and how the type of decision problem moderates this effect. Informed by the theoretical analysis, we conducted four studies designed to test this effect with high power. None of the studies produced significant effects of switch rate. Lastly, the studies replicated a prior finding that optional stopping and switch rate are negatively correlated. In sum, this research elaborates a fuller understanding of the relation between search strategies (switch rate and optional stopping) on how people decide which option is better and their tendency to overweight versus underweight rare outcomes.

KEYWORDS

decisions from experience, optional stopping, replication, sampling, search strategies, underweighting

1 | INTRODUCTION

Imagine a customer deciding which of two restaurants is better. When determining how to sample the two restaurants, there are multiple choices that must be made. First, the customer could decide to dine at each restaurant a certain number of times (e.g., five times each) before making a decision, known as a "fixed sampling strategy." Alternatively, the customer could decide to try the restaurants without a predetermined number of dining experiences and to stop sampling whenever they feel like they have enough evidence that one is better than the other, known as an "optional-stopping strategy." Second, the customer could use different patterns of sampling the restaurants. She could dine at Restaurant A multiple times followed by Restaurant B multiple times. Alternatively, she could switch back and forth between the two restaurants or use a sequence between these two extremes. In the current research, we investigated how these two features of sampling, fixed versus optional stopping, and the switch rate, influence peoples' final decisions, especially in the context of rare but extreme outcomes.

Although the relations between switch rate and optional stopping on participants' final decisions have been studied previously, there are some critical open questions. With regard to switch rate, Hills and Hertwig (2010; hereafter HH) found that people who switched more frequently when sampling made choices as if they were underweighting rare outcomes. However, only 6% of HH's data could be analyzed, resulting in a very small sample; we conducted four studies to test this effect with a larger sample and under a variety of paradigms.

With regard to the relation between optional stopping and peoples' final decisions, we examined an intuitive but novel hypothesis; if people stop sampling after a rare but extreme outcome (e.g., stop testing the restaurants after getting food poisoning from one), the optional stopping could make the rare outcome seem more common than it actually is, which could bias peoples' final decisions.

In the remainder of the introduction, we first broadly review research on the sampling strategies and decision policies people use when making decisions from experience. Next, we review research on the relation between switch rate and final decisions and identify questions for future studies. Afterwards, we review research and discuss open questions on the relation between optional stopping and final decisions. Finally, we outline four studies on switch rate and optional stopping.

2 | RESEARCH ON SAMPLING STRATEGIES IN DECISIONS FROM EXPERIENCE

2.1 | Introduction to decisions from experience (DFE)

In the *decisions from experience* paradigm (DFE; Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009), participants learn the possible outcomes associated with the different options and the probabilities of each outcome from experience. It is similar to the restaurant example above, although typically involves monetary gambles instead of restaurant experiences. The DFE paradigm consists of two stages. In the *sampling phase*, participants sample from the two options and view the associated monetary outcomes. Participants are free to choose which option to sample at each opportunity without incurring any cost or obtaining any reward (for a discussion of related versions of the task, see Gonzalez & Dutt, 2011; Hills & Hertwig, 2012; Lejarraga, Hertwig, & Gonzalez, 2012; Mehlhorn et al., 2015).

For example, in one prototypical case, which we label Problem 1, when a participant chooses to sample Option A, they receive a reward of 32 points with a 10% probability and 0 points with a 90% probability. When a participant chooses to sample Option B, they always receive 3 points. Because participants are typically free to choose when to stop sampling, it is possible that a participant will stop sampling before experiencing the 32-point outcome; participants never know for sure whether they have experienced all the possible outcomes.

After participants feel like they can judge the value of the options, they stop sampling and enter the *choice phase*. In this phase, participants make one choice between the two options. The outcome that they experience counts towards the total reward that they obtain at the end of the study.

2.2 | Introduction to decision policies in DFE

Researchers have investigated numerous *decision policies*—how participants decide which option is better for the final consequential choice (Erev et al., 2010; Hau, Pleskac, Kiefer, & Hertwig, 2008). The main decision policy, sometimes called the *summary* policy or natural mean heuristic, is to select the option with the higher mean outcome of the experienced sample (Hau et al., 2008; Hertwig & Pleskac, 2010; see also Fox & Hadar, 2006; Ungemach, Chater, & Stewart, 2009). This policy approximates the long-run expected value (EV) using a limited sample. In contrast, there are other decision policies based on heuristics and reinforcement learning that do not approximate EV (Hau et al., 2008; Thorngate, 1980; Weber, Shafir, & Blais, 2004).

Relevant to the current study, HH proposed another heuristic called the *round-wise* policy, which involves comparing sequential pairs of samples. Suppose a participant chooses the following samples in order (the outcome for each sample appears in superscript): $A^0B^3A^0B^3A^0B^3A^0B^3A^0B^3A^0B^3A^{32}B^3A^0B^3A^0B^3A^0B^3$. The round-wise policy involves comparing the first sample of A^0 with the first sample B^3, the first sample B^3 with the second sample A^0, and so forth. For each comparison, called a "round," the option with the higher outcome wins, and the option that wins more round-wise comparisons is preferred. Because B almost always has a higher value than adjacent samples of A, the round-wise policy concludes that B is better. In contrast, the summary policy concludes that A is better because $EV_A = 3.2$ and $EV_B = 3.0$.

2.3 | Moderators of apparent underweighting versus overweighting of rare outcomes in DFE

Much research has studied the accuracy of peoples' decisions from experience. Most of it has focused on situations in which one option has a rare outcome (e.g., the rare 32-point outcome for Option A), and whether and when people over or underweight the rare outcome. The logic is as follows: Using Problem 1 as an example, if people tend to choose Option B (a sure bet of 3 points), it is said that they are acting "as if" they are underweighting the 10% chance of obtaining 32.[1] A number of factors moderate apparent overweighting versus underweighting of rare outcomes.

First, much of the research has focused on comparing how participants perform in the DFE paradigm, to the *decisions from description* (DFD) paradigm. In the DFD paradigm, participants are told the exact probabilities of prospects rather than learn them via sampling. In the DFD paradigm, people tend to overweight rare events (Barron & Erev, 2003; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Weber et al., 2004). In contrast, in the DFE paradigm, people appear to act as if they underweight rare events (Hertwig et al., 2004; Hertwig & Erev, 2009; Wulff, Mergenthaler-Canseco, & Hertwig, 2018, but see Noguchi & Hills, 2015, when there are more than two options). The "description-experience gap" may be due in part to the different presentation formats. In DFE, outcomes are represented as natural frequencies rather than probabilities, which may trigger the use of different algorithms for evaluating the information (Hau et al., 2008; Hau, Pleskac, & Hertwig, 2010), though the evidence is mixed (see Rakow, Demes, & Newell, 2008).

[1] In line with previous authors, we use the phrases "apparent underweighting" or "act as if they were underweighting" because assessing underweighting versus overweighting is tricky. Participants do not know the true probabilities of all the outcomes; they may not even have experienced the rare 32-point outcome. From the participants' perspective, there might even be other rare outcomes that they have not experienced.

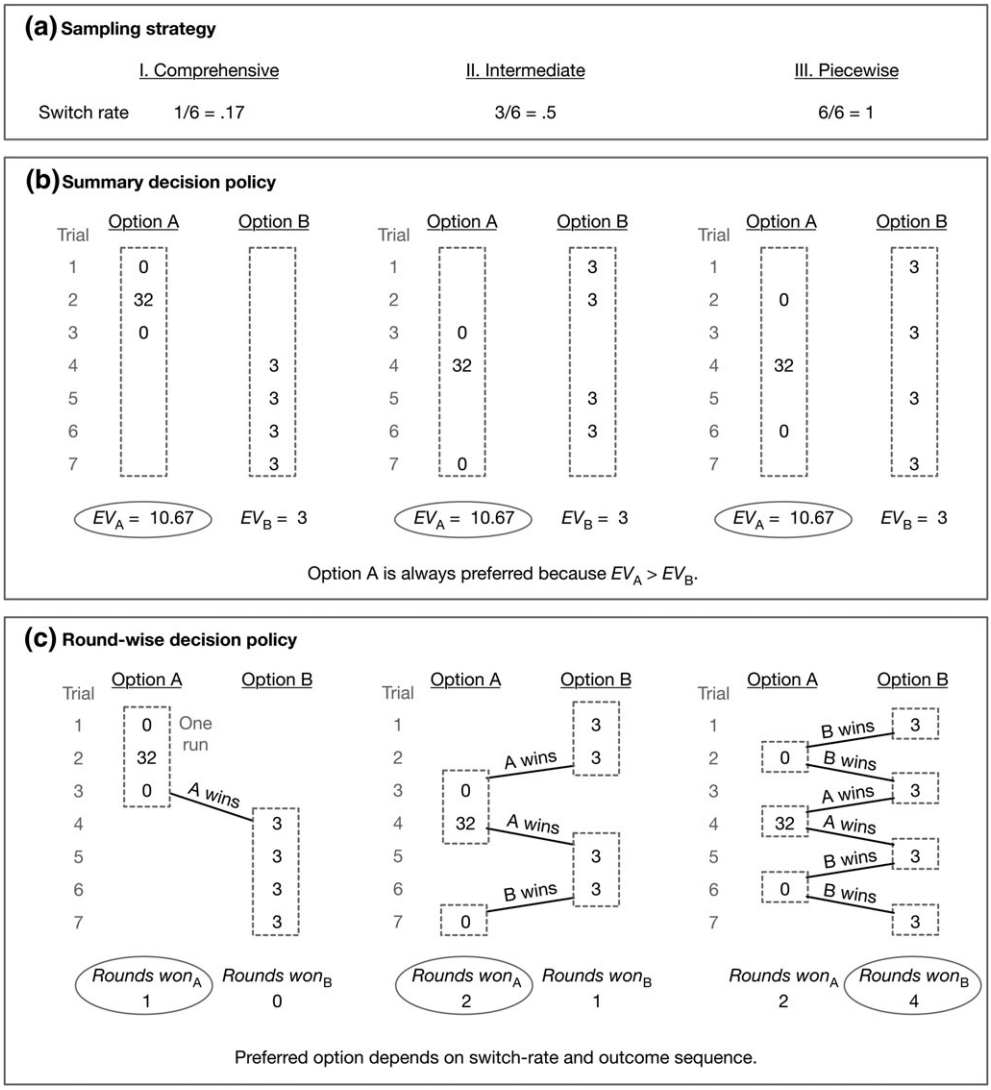

FIGURE 1 (a) Three sampling strategies with low, intermediate, and maximum switch rates. Possible outcome sequences when using a given strategy are shown in the following two panels, along with how each sequence is evaluated using a particular decision policy. (b) Implementation of a summary decision policy. Expected value (EV) is computed for all outcomes within the dashed gray boxes. To use the summary policy, the EV of all outcomes from each option is computed. The option with the higher EV is preferred for the choice phase (circles). (c) Implementation of the round-wise decision policy. Each gray box denotes one run (see left outcome sequence). Each switch between runs creates a round (indicated by the solid black lines between runs). Different sampling strategies partition the outcomes into rounds differently. Each round is won by the option that has a higher EV. The option that wins more rounds is preferred for the choice phase (circles)

Second, underweighting is in part the result of sampling error (the experienced probabilities do not match the stated probabilities); sampling error is greater with smaller samples. People underweight less if they are incentivized to draw larger samples (Hau et al., 2008; Hertwig & Pleskac, 2010).

Third, underweighting is moderated by the structure of the decision problems. Wulff et al. (2018) found that underweighting is greater when the rare outcome is extremely rare, when one option is risky (has more than one possible outcome) while the other is safe (has only one outcome), and when the prospects contain losses.

The current research focuses on two other factors potentially related to apparent overweighting and underweighting of rare outcomes; participants' switch rates during sampling and optional stopping behavior. The next section offers an overview of prior work on switch rate and explains our motivation for further theoretical and empirical research. The subsequent section covers our motivation for studying stopping behavior.

3 | THE RELATION BETWEEN SWITCH RATE AND FINAL DECISIONS

3.1 | Prior research on sampling switch rate

Participants' switch rates in DFE tasks are influenced by a number of factors. First, age and numeracy are negatively related with switch rate (Wegier & Spaniol, 2013). Second, the switch rate can also be influenced by structural properties of the decision problems; participants switch more when there are more options (Hills, Noguchi, & Gibbert, 2013; Noguchi & Hills, 2015, 2016). Third, during the sampling phase, there is a general shift from a high switch rate ("exploration") to a lower

switch rate ("exploitation"—choosing the option thought to be better), even when participants only obtain a reward from the final consequential choice and not during the sampling phase (Gonzalez & Dutt, 2012, 2016; Gonzalez, Lerch, & Lebiere, 2003; Lejarraga, Dutt, & Gonzalez, 2010); however, see Hills and Hertwig (2012) for controversy over this finding.

There has been some research on the link between sampling strategy and subjects' final choices. Rottman (2016; random order condition) found that there was typically no relation between the switch rate and the accuracy of the final judgment. In contrast, Wegier and Spaniol found that participants were more accurate in choosing the option with the higher average value when they used a higher switch rate (Wegier & Spaniol, 2013; Wegier & Spaniol, 2014a; Wegier & Spaniol, 2014b; see also Wegier, Bianchi, & Spaniol, 2015). However, the learning problems used in these studies were fairly different from the distributions in the classic "gamble" problems used in the DFE paradigm.

3.2 | HH's proposed link between switch rate, decision policies, and underweighting rare outcomes

Hills and Hertwig (2010) proposed that individuals who switch more frequently between options are more likely to use the round-wise instead of the summary decision policy and subsequently more likely to underweight rare outcomes. The intuitive reason is that switching draws attention to the comparison between the rounds. Whether or not this actually occurs is one of the two central questions of this paper. The rest of this section explores the details of and presents a new theoretical analysis of HH's theory.

Figure 1a depicts three sampling strategies an individual could use. Each strategy tests Option A three times and Option B four times but in different orders. The critical difference is the switch rate—the number of switches between the two options divided by the number of possible switches (total samples minus one).

Figure 1 also depicts the two decision policies. The summary policy (Figure 1b) is not sensitive to the sampling strategy, so Option A is preferred regardless of the switch rate. In contrast, because the round-wise policy (Figure 1c) is influenced by how outcomes are grouped into rounds,[2] it is sensitive to the switch rate. In Figure 1c, A wins more rounds for the switch rates of .17 and .5, but B wins more rounds for the switch rate of 1.

There are three important points to note about this theory. First, this theory predicts that underweighting of rare outcomes will occur under higher switch rates (e.g., rightmost example in Figure 1c). Under the round-wise policy, rare events only affect rounds in which they appear, and higher switch rates produce more rounds, dulling the effect of rare extreme outcomes. Second, at very high switch rates, the two decision policies can diverge (Figure 1) or converge. For example, in the following sequence, both policies choose B: $A^0B^3A^0B^3A^0B^3A^0B^3A^0B^3A^{32}B^3A^0B^3A^0B^3A^0B^3A^0B^3$. Third, when a participant only switches once, as in the leftmost column of Figure 1, both the round-wise and summary policies make the same predictions.

3.3 | Switch rates determine when decision policies diverge

The analysis above only presents a qualitative overview of the relations between switch rates, decision policies, and underweighting. Higher switch rates are believed to be correlated with use of the round-wise (as opposed to summary) policy empirically, which is in turn believed to mathematically lead to underweighting rare outcomes. We conducted simulations to better understand this aspect of HH's theory. Most studies on DFE have used subsets of the 13 problems listed in Table 1,[3] so we used these for the simulations. Each of the 13 decision problems was simulated with 40 samples, and with switch rates of 2, 6, 10, 14, 18, 22, 26, 30, 34, and 38 times out of 39 possible switches, 1,000 times each. Figure 2 plots the probability of divergence by switch rate.

The first critical finding is that when the round-wise and summary policies diverged (the summary and round-wise policies predict opposite choices,[4] e.g., the rightmost column of Figure 1), the round-wise policy chose the option consistent with "underweighting" 97.6% of the time. For this reason, as well as difficulties defining underweighting,[5] we focused our simulations on when divergence is likely to occur.

[2]In our analysis, each run is compared to both the prior and subsequent runs of the other option. For example, in the sequence $A^0A^0B^3B^3B^3A^{32}A^0$, the A^0A^0 run is compared against the $B^3B^3B^3$ run, and the $B^3B^3B^3$ run is compared against the $A^{32}A^0$ run. In contrast, HH first grouped the runs into pairs and compared the two runs within each pair; the A^0A^0 run would be compared with the $B^3B^3B^3$ run, but the $A^{32}A^0$ run would be ignored. These two definitions frequently result in the same outcomes, and an analysis of HH's data using our method found the same pattern of results. However, we did not use HH's method because it ignores the last run if the number of runs is odd, and arbitrarily compares certain pairs of runs but not other adjacent pairs.

[3]Of the studies analyzed by HH, Hau et al. (2008), Hertwig et al. (2004), and Ungemach et al. (2009) used Problems 1, 2, 4, 9, 10, and 12. The remaining problems in were introduced by Hertwig and Pleskac (2010).

[4]Cases in which one policy predicts a tie for A and B, and the other policy prefers one option over the other were not considered as divergent. In the analyses in our studies, including these cases did not lead to differences in results.

[5]For Problem 1 in Table 1, Option B is considered to be consistent with underweighting the 10% probability of 32, because a decision maker maximizing EV should choose B so long as they believe that 32 has a probability no more than 9.375%. In fact, it has a probability of 10%, so choosing Option B is consistent with believing that it has a lower probability than it actually does. For Problem 2, it is similar, but the opposite because the decision maker should choose the option that results in less of a negative outcome. For Problem 9, Option B is considered to be consistent with overweighting the 20% chance of 0, because a decision maker maximizing EV should choose B only if they believe that the probability of 0 is greater than 25%. Since B is considered "consistent with overweighting," by process of elimination, A is considered "consistent with underweighting" (even though a rational decision maker would choose B if they accurately thought that 0 has a probability of exactly 20%). There are two issues with this definition. First, in past studies, rare outcomes were defined as those appearing with a probability of .2 or less, but Problems 12 and 13 have outcomes that occur with a probability of .25, highlighting the arbitrariness of this cutoff (Hertwig et al., 2004, Footnote 2). Second, in prior research, the options marked ^{U} were considered to imply "underweighting" even if that option happened to coincide with the higher expected value. For example, suppose that for Problem 1, the outcome of 32 occurred less than 10% of the time, such that the mean outcome for B was higher than the mean outcome of A. Choosing Option B was still considered underweighting the rare outcome even though the correct option was chosen on the basis of the expected value (see Appendix A for more details). For consistency with the prior research, we adopted the same definition of underweighting.

TABLE 1 Decision problems used in the decisions from experience task

Problem	Higher EV	Option A		Option B		Used in studies			
		Outcome	Pr.	Outcome	Pr.	1	2	3	4
Group I: Maximal divergence									
1	A	32	.1	3^{U}	1	✓	✓	✓	✓
2	A	−3	1	-32^{U}	.1	✓			✓
3	A	16	.2	3^{U}	1				✓
Group II: Moderate divergence									
4	A	32	.025	3^{U}	.25	✓			✓
5	A = B	10	.1	1^{U}	1				✓
6	A = B	10^{U}	.9	9	1			✓	✓
7	A = B	−10	.9	-9^{U}	1				✓
8	A = B	10	.05	1^{U}	.5				✓
Group III: A little divergence									
9	A	4^{U}	.8	3	1	✓			✓
10	A	-3^{U}	1	−4	.8	✓			✓
Group IV: Almost no divergence									
11	B	32	.025	3^{U}	1				✓
Group V: Decreasing divergence									
12	A	4	.2	3^{U}	.25	✓			✓
13	A	−3	.25	-4^{U}	.2				✓

Note. For all probabilistic outcomes, the other unmentioned probability is an outcome of 0. ^{U} indicates option defined as "underweighting" the rare outcome.

Figure 2 reveals two further insights. First, the probability of divergence increases with the switch rate for all problems except for Problems 12 and 13. Second, the problems cluster into five groups.

In the Group I and II problems, the rate of divergence increases with the switch rate and asymptotes at either slightly above 50% (Group I) or right around 50% (Group II). The reason is that the round-wise policy almost always prefers B. For Group I, the summary policy prefers A at rates slightly higher than 50% because EV_A is only slightly higher than EV_B. For Group II, EV for the two options are almost, if not exactly, identical.

In Group III, the divergence is lower because Option A usually wins according to both policies. The single problem in Group IV has very low divergence for all switch rates; both policies almost always prefer Option B.

The problems in Group V lead to increasing and then decreasing rates of divergence. Both options A and B have rare outcomes of similar probabilities, which means that the round-wise policy frequently produces a tie (which does not count as divergence). When the round-wise policy does not produce a tie, the summary and round-wise policies usually converge.

Practically, one important insight from the simulation was identifying that Groups I and II are most likely to lead to divergence; we designed our studies to focus on problems in these groups. Another important insight is that problems with the highest rates of divergence (Groups I and II) have very similar EVs for A and B. This, combined with the fact that divergence is low in general, means that the round-wise policy actually does a fairly good job of identifying the better option, and when it does choose the worse option, it is not all that much worse. Thus, the round-wise policy could be considered a good heuristic.

3.4 | HH's theory on switch rate: Empirical evidence and motivation for the current research

HH investigated the link between switch rate and decisions in datasets from four prior studies (Experiments 1 and 2 of Hau et al., 2008; Hertwig et al., 2004; Hertwig & Pleskac, 2010; Ungemach et al., 2009). The main finding was that in the subset of scenarios for which the summary and round-wise policies diverged, participants who had a higher (lower) switch rate were more likely to make a choice consistent with the round-wise (summary) policy.

HH also found a correlation between switch rate and underweighting rare outcomes. This analysis only included scenarios for which the rare outcome was experienced at least once. However, an error in that analysis was recently uncovered, prompting a correction stating that the correlation was not significant (Hills & Hertwig, 2017; see Appendix A for another analysis reported in the correction including all scenarios). The current research was conducted before this correction was produced. In light of the correction, and given the aforementioned limitations of the underweighting analysis, we consider the analysis of divergence to be more informative. Still, we report the analysis of underweighting, for three reasons: underweighting is theoretically tied to switch rate, the low rates of divergence reduces power for the decision policy analysis, and because certain extreme manipulations (Study 1) can only be studied with the underweighting analysis.

We sought to gather additional evidence of the relation between switch rate and decision policies for several reasons. First, HH's analysis involved a small subset of the total data. HH started

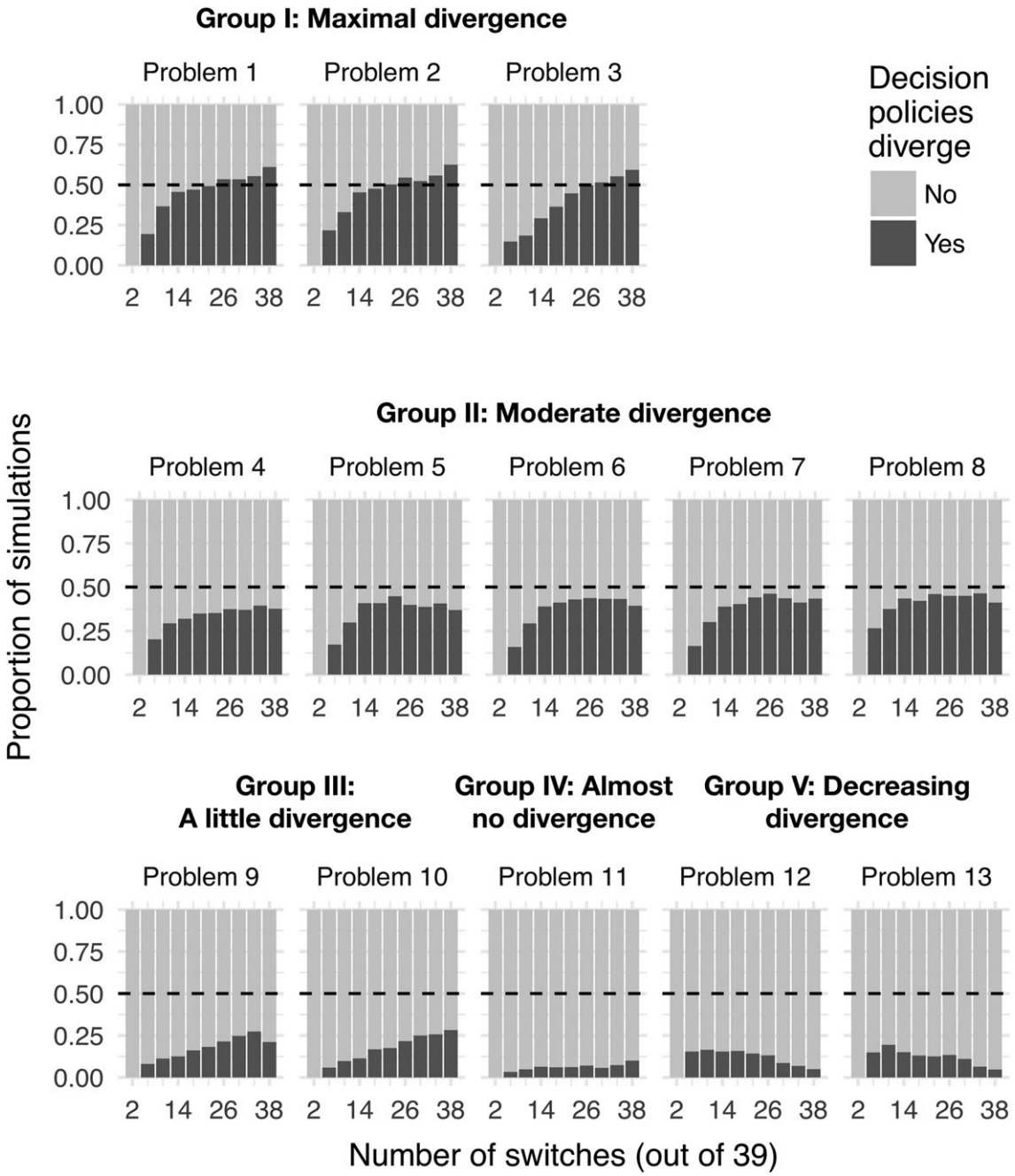

FIGURE 2 Proportion of simulations for which the summary and round-wise policies diverge at different switch rates. Proportions calculated for 1,000 simulations at each switch rate for each problem. Sampling phases in these simulations were fixed at 40 samples

with 1,818 scenarios and discarded scenarios in which rare outcomes were never experienced and in which divergence did not occur, leaving only 104 scenarios to analyze.[6]

Second, HH's data were obtained from a "free-sampling" paradigm in which participants chose which options to sample, and how many total samples to collect. There are a couple benefits of conducting a "forced-sampling" study instead. Forcing participants to sample options in a particular sequence allowed us to ensure the probabilities they experienced matched the "true" probabilities they were supposed to experience in the long run, circumventing the problem of participants not experiencing rare outcomes. Second, randomizing participants to sample with a high versus low switch rate allowed us to test the causal influence of switch rate on underweighting and decision policy usage. Third, the forced sampling paradigm allowed us to implement a very strong manipulation of switch rate which should, in theory, result in a stronger effect. In HH, there was a wide range of switching rates, though fairly few greater than .50.

Third, HH used a subject-level median split to classify frequent versus infrequent switchers. This analysis reduces power and ignores the possibility that participants could have used different sampling strategies in different scenarios. We used hierarchical logistic regression to analyze the effect of switch rate on decisions within a given

[6]These data come from the correction to HH (Hills & Hertwig, 2017). The subset was based on our definitions of the *round-wise* policy and of divergence.

scenario, while controlling for repeated measures and clustering within participants and problems.[7]

We conducted four studies to test whether a higher switch rate is related to underweighting and using the round-wise decision policy. In Studies 1 and 2, to increase power, we chose problems with high rates of divergence, ensured that participants actually saw the rare outcomes and manipulated the switch rates to be very low versus very high. In Studies 3 and 4, we allowed subjects to sample freely to test if free sampling was necessary for the effect.

4 | THE RELATION BETWEEN OPTIONAL STOPPING AND FINAL DECISIONS

At a high level, this research examines the roles of two features of sampling behavior, switch rate and optional stopping, on participants' final decisions. We now focus on the second question regarding optional stopping. Optional stopping refers to situations in which participants decide how much data to collect in a sequential manner, which can be influenced by many factors (Fried & Peterson, 1969; Gonzalez & Dutt, 2011, 2016; Markant, Pleskac, Diederich, Pachur, & Hertwig, 2015; Wulff et al., 2018).[8]

One reason to attend to optional stopping is that previous studies have found that participants who use a higher switch rate tend to stop sampling earlier (Hills & Hertwig, 2012; Rakow et al., 2008); all else being equal, fewer samples could mean worse decisions. Thus, when investigating switch rate on final decisions, it is also important to consider the role of optional stopping.

We considered a particular form of optional stopping; optional stopping after rare outcomes, which we term "rare outcome stopping" or RO-stopping for short. Suppose a participant chooses the following samples and experiences the corresponding outcomes in Problem 1: $B^3B^3B^3A^0B^3A^0B^3A^{32}$. The outcome of 32 may be so tempting that the participant stops sampling and chooses Option A for the final decision. Likewise, in the negative domain (e.g., Problem 2: $A^{-3}B^0A^{-3}B^0A^{-3}B^0A^{-3}B^{-32}$), a participant might stop sampling after a large negative outcome, akin to the "hot-stove" effect (Denrell, 2007; Plonsky & Erev, 2017).

We predicted that RO-stopping would lead to decisions consistent with overweighting the rare outcome. In Problem 1, experiencing A^{32} and then selecting Option A is consistent with overweighting the rare 32 outcome. In Problem 2, experiencing B^{-32} and then selecting Option A is also consistent with overweighting the −32 outcome (see Table 1). In contrast, we predicted that when participants stop sampling after experiencing a "common" outcome (e.g., B^3 or A^0 in Problem 1), they would be relatively more likely to choose the option that underweights the rare outcome (B).

We also predicted that RO-stopping would lead to decisions relatively more consistent with the summary (as opposed to the round-wise) decision policy. Consider the $B^3A^0B^3A^0B^3A^0B^3A^{32}$ sequence. After the penultimate choice (B^3), both the summary and round-wise decision policies prefer B. After the A^{32} experience, the round-wise policy still prefers B (B won 6 out of 7 rounds), but the summary policy prefers A (EV_A = 8 and EV_B = 3). Using the summary policy in combination with stopping after a rare event should not be viewed as optimal because it greatly exaggerates the difference in EV between the two options. In fact, one might argue that the round-wise decision policy is a useful heuristic in that it could mitigate oversensitivity to extreme outcomes.

In sum, we investigated a novel hypothesis that when participants stop sampling immediately after experiencing a rare outcome, they would act as if they *overweight* the rare outcome, which is also consistent with using a summary policy instead of a round-wise policy.

5 | OUTLINE OF STUDIES

We conducted four studies to investigate the questions about how switch rate and optional stopping influence participants' final decisions. In Study 1, the goal was to study the role of switch rate on final decisions using a strong manipulation of switch rate and prohibiting optional stopping, thereby eliminating it as a complicating factor. Study 2 was very similar; the switch rate and total number of samples were fixed; however, participants had some flexibility of *when* (not how much) to switch. The goal was to see if active involvement in switching had an influence on final decisions. In Studies 3 and 4, participants controlled both the switch rate and the total number of samples, which allowed for an analysis of both switch rate and optional stopping. Study 3 was conducted online, whereas Study 4 was conducted in the lab.

To presage the results, we found no evidence of the link between switch rate and participants' final decisions, either in terms of underweighting or making choices in line with the summary versus round-wise decision policies. However, we found that when participants engage in RO-stopping, they were more likely to make final decisions consistent with apparent overweighting and consistent with the summary policy instead of the round-wise policy.

6 | STUDY 1: FORCED SAMPLING

In Study 1, participants either sampled one option repeatedly before switching once and sampling the other option repeatedly or switched between options on each trial. We began our attempt to replicate the findings by Hills and Hertwig (2010) with this extreme manipulation because these sampling strategies represent the two theoretical extremes they discussed. HH predicted that a higher switch rate is more likely to result in choosing the option that underweights the rare outcome. The design of this study only permitted an analysis of the relation between switch rate and underweighting, not decision policy, because the manipulation was so strong.[9]

[7] An analysis of HH's data using hierarchical regression revealed the same effects that HH reported. Specifically, the analysis of the relation between switch rate and decision policies was significant, and the analysis of the relation between switch rate and underweighting was not (see Tables 2A and 2B). Using a scenario-level median split to classify scenarios as having high versus low switch rates reveals the same qualitative pattern of results.

[8] We thank Ralph Hertwig and an anonymous reviewer for suggesting that we investigate optional stopping.

[9] The summary and round-wise policies always converged for participants who switched only once (low switch rate) sampling strategy (see Figure 1). This made it impossible to test if switch rate influences whether decisions are consistent with one policy or the other.

6.1 | Method

6.1.1 | Participants and payment

One hundred participants were recruited from Amazon MTurk for a base payment of $1.50. The study lasted eight min on average. One additional participant completed one scenario before withdrawing; their data were included in the analyses. The goal of the DFE task is to win points from the choice phase of each decision problem. At the end of the experiment, the points that participants earned (Table 1) were converted into a monetary bonus; 1 point equaled one cent, the same payment structure used by Hertwig and Pleskac (2010). Because some of the decision problems had negative outcomes (Table 1), participants began the experiment with 50 points to ensure that they would end with a positive bonus. On average, participants received a bonus of $0.51.

6.1.2 | Design and procedure

Participants completed Problems 1, 2, 4, 9, 10, and 12, the original problems used in Hertwig et al. (2004), in a random order.

In each problem, participants drew 40 samples in the sampling phase, one at a time, comprising 20 samples of Option A and 20 samples of Option B. The outcomes associated with the two options followed the probabilities in Table 1 exactly; rather than being drawn at random for each sample, the frequency of each outcome were predetermined and the trial order randomized within an option. For Problem 4, the .025 probability was converted into .05.

Half the participants were forced to switch only once; 20 samples from one option followed by 20 samples from the other. The other half alternated between the two options. Participants chose whether to start with the left or right option on screen and the two options for each problem in Table 1 were randomly mapped onto the two positions on the screen. After clicking a button to sample the option, the outcome appeared for 1 s, after which participants could move to the next sample. The outcomes during the sampling phase did not count towards the point total.

After receiving the 40 samples, participants proceeded to the choice phase; they made one choice between the two options and gained those points towards their overall point total. The outcome for this choice was determined randomly on the basis of the probabilities in Table 1. Participants then advanced to the following problem until they completed all six problems. Throughout the study, the total points that they had earned was displayed on the screen, and at the end of the study, participants were paid a monetary reward based on their point total.

6.2 | Results

We analyzed whether participants who used a high switch rate were more likely to underweight rare outcomes. Because the experienced data perfectly matched the probabilities in Table 1, all the rare outcomes were experienced, so no data were discarded.

We ran a logistic regression predicting the likelihood of underweighting rare outcomes based on switch rate. Due to the repeated measures (each participant experienced all six problems), the regression included random crossed by-subject and by-problem intercepts and slopes for switch rate. The maximal model failed to converge, so we dropped the correlation parameters between the random intercepts and slopes, as recommended by Barr, Levy, Scheepers, and Tily (2013). In all regression analyses across all our studies, whenever the maximal models failed to converge, we dropped parameters according to recommendations by Barr et al. (2013).

The probability of apparent underweighting was .46 in the low switch rate condition and was .51 in the high switch rate condition. Unlike in HH, the regression did not reveal a significant effect of switch rate on choosing options that underweighted the rare outcome, $B = 0.17$, CI [−0.22, 0.57], $p = .39$. For comparison with other studies, the regression weight represents the difference of switch rate from 0 to 1, though in reality the switch rates were 1/39 versus 39/39. Table 2A summarizes the regression results of underweighting for all four studies.

TABLE 2A Logistic regression results of switch rate predicting choices that underweight the rare outcome

		Pr. of underweighting the rare outcome and 95% CI at different switch rates			Log odds		
Study	Switches	Low	Medium	High	*B*	95% CI	*p*
Study 1	1 vs. 39	.46 [.36, .59]		.51 [.40, .63]	.17	[−.22, .57]	.39
Study 2	3 vs. 20 vs. 37	.57 [.49, .64	.52 [.42, .62]	.54 [.45, .63]	−.14	[−.70, .42]	.63
Study 3	Free sampling	.41		.41	−.001	[−.92, .92]	.99
Study 4	Free sampling	.47		.45	−.10	[−.71, .51]	.75
Hills and Hertwig (2010)	Free sampling	.48		.53	.21	[−.39, .81]	.50

TABLE 2B Logistic regression results of switch rate predicting choices consistent with the round-wise decision policy

		Pr. of round-wise decision policy and 95% CI at different switch rates			Log odds		
Study	Switches	Low	Medium	High	*B*	95% CI	*p*
Study 2	3 vs. 20 vs. 37	.57 [.44, .69]	.38 [.26, .51]	.43 [.31, .55]	−.63	[−1.46, .18]	.13
Study 3	Free sampling	.34		.46	.51	[−.35, 1.36]	.25
Study 4	Free sampling	.31		.38	.29	[−.70, 1.27]	.57
Hills and Hertwig (2010)	Free sampling	.35		.77	1.83	[.27, 3.38]	.02

Note. For Studies 1 and 2, the probabilities and 95% CIs are presented for each experimental group. For Studies 3, 4, and Hills and Hertwig (2010), we report the predicted probabilities for switch rates of 0 (low) and 1 (high).

6.3 | Discussion

Higher switch rates were unrelated to underweighting rare outcomes. Unlike the studies analyzed by HH, participants in Study 1 had no control over their sampling strategies and the number of samples drawn from each option. Perhaps a better test of HH's theory requires a task that more closely resembles the free-sampling paradigm used in the studies analyzed in HH. Research on various tasks suggests that different cognitive processes are involved when people sample information actively versus passively (e.g., Lagnado & Sloman, 2004; Markant & Gureckis, 2014; Wulff et al., 2018). In Study 2, we gave participants a little more control over their sampling, to see if this was a necessary condition for their sampling strategies to influence their final choices.

7 | STUDY 2: NEARLY FORCED SAMPLING

In Study 2, we gave participants some degree of control over their sampling but still ensured large differences between the high and low switch rate conditions. Furthermore, unlike in Study 1, in Study 2, the summary and round-wise policies sometimes diverged even at low switch rates, allowing us to also analyze whether more frequent switching leads to more use of the round-wise policy.

7.1 | Method

7.1.1 | Participants and payment

Five hundred participants were recruited from MTurk for a base payment of $0.50. The study lasted 2 min on average. Participants began the experiment with zero points. A bonus of 3 cents was paid to 64.2% of participants; a bonus of 32 cents was paid to 4.4% of participants; and the remaining participants were paid no bonus.

7.1.2 | Design and procedure

Only Problem 1 was used because it comes from Group I, which has the highest rate of divergence (Figure 1), allowing us to discard the fewest number of scenarios. Each participant worked with just one scenario. Unlike Study 1, since participants could choose how many samples to draw from each option, the outcomes were determined probabilistically.

All participants had to draw a total of 40 samples. There were three conditions: participants made 3, 20, or 37 switches, corresponding to a low, medium, or high switch rate, respectively. Participants decided when they would switch within a given problem. Participants were told the required number of switches in advance and were allowed to sample freely until (a) they made the required number of switches, after which only the previously sampled option could be sampled, or (b) the number of remaining trials equaled the number of switches remaining, after which they were forced to switch options for all the remaining draws. After each draw, the number of switches and samples they had remaining were displayed. After 40 samples, participants made their final choice.

Based on simulations, the probability of divergence with only three switches was about half as likely as in the other two conditions. In order to increase the number of scenarios with divergence, 250 participants were assigned to the condition with three switches, and 125 were assigned to each of the other conditions.

7.2 | Results

Following HH, we first eliminated the scenarios in which the rare outcomes never occurred; 361 scenarios remained (72.2% of all 500 scenarios), which were used for analyzing the relation between switch rate and underweighting the rare outcome. There were 157, 93, and 111 scenarios in the low, medium, and high switch rate conditions, respectively.

7.2.1 | Switch rate and underweighting

The probability of apparent underweighting was .57, .52, and .54 for the low, medium, and high switch rate conditions, respectively (Table 2A). A logistic regression using switch rate (3/39, 20/39, and 37/39) as the predictor was not significant, $B = -0.14$, CI [−0.70, 0.92], $p = .63$. (The regression weight reported is for the estimated difference of switch rates of 0 versus 1.)

7.2.2 | Switch rate and decision policies

To analyze the relation between switch rate and decision policy, we analyzed scenarios for which the summary and round-wise policies diverged (182 scenarios; 36.4% of all 500 scenarios). There were 56, 58, and 68 scenarios in the low, medium, and high switch rate conditions, respectively.

Across the low, medium, and high switch rate conditions, the probability of making a decision consistent with the round-wise decision policy was .57, .38, and .43, respectively (see Table 2B for a summary of the decision policy results for all studies). This difference was not significant according to a logistic regression, $B = -0.63$, CI [−1.46, 0.18], $p = .13$.

The trend among the three conditions was not monotonic, and in fact, the low-switch rate condition had the highest probability of making a decision consistent with the round-wise policy, which goes against HH's prediction.

7.3 | Discussion

Again, we did not find a significant relation between switch rate and underweighting. Additionally, we found that switch rate was unrelated to decision policy.

8 | STUDY 3: FREE SAMPLING WITH AN MTURK SAMPLE

Perhaps the effect proposed by HH that switch rate influences underweighting and decision policies, only holds if participants actually have a high level of control over the sampling process; in the studies evaluated by HH, participants had complete control over how to test the two options. Whether or not participants have control over which options to sample and when to stop can influence their final choices in other ways (Hertwig & Pleskac, 2010; Rakow et al., 2008; Wulff et al., 2018). Studies 3 and 4 were designed to test whether the relations between switch rate with underweighting and decision policies hold when participants are in control of sampling.

In addition, Studies 3 and 4 also allowed us to test the RO-stopping hypothesis outlined in the introduction. This hypothesis predicts that participants who stop sampling after experiencing a rare outcome would tend to overweight the rare outcome, and their decisions would be more aligned with the summary than round-wise decision policy.

In order to maximize the number of usable scenarios, our participants worked with Problems 1 and 6. Our simulation showed that these two decision problems have relatively high rates of divergence, and we chose problems from different groups for diversity. To ensure that we would have enough scenarios in which the decision policies diverged, we used a larger sample than in the prior studies.

8.1 | Method

8.1.1 | Participants and payment

Eight hundred participants were recruited from MTurk for a base payment of $0.50. Twelve additional participants completed the study but did not claim payment, and one further participant completed one of the two problems; their data were included in the analyses. The study lasted 4 min on average. At the end of the study, participants were paid an additional one cent for each point they obtained, and participants began the study with zero points. On average, participants received a bonus of $0.12.

8.1.2 | Design and procedure

The procedure mimicked the free sampling task of previous studies (e.g., Hau et al., 2008; Hertwig et al., 2004). In the sampling phase, participants could choose the sampling order and total samples to obtain from each of the two options. Participants could only advance to the choice phase after sampling at least twice from each option. Participants worked with Problems 1 and 6 in a counterbalanced order.

8.2 | Results

For all the results below, we eliminated scenarios in which rare outcomes never occurred, leaving 574 scenarios, 35% of the original 1,625 scenarios. This was the subset used for the analysis of underweighting. For the analysis of decision policies, a further subset of the scenarios was taken in which the summary and round-wise policies made diverging predictions, consisting of 258 scenarios (16% of the original total).

8.2.1 | Sampling phase

A histogram of the number of samples drawn and the switch rate is displayed in Figure 3. These histograms present the larger sample of 574 used for the analysis of underweighting in lighter gray, and the smaller sample of 258 used for the decision policy analysis (see below) in darker gray.

With regard to the total samples, the two subsets look fairly similar. However, in the subset of 574 scenarios in which all outcomes occurred, participants drew fewer samples in the present study (M = 16.95, SD = 15.28) than in HH (M = 31.39, SD = 25.59). The switch rates in that same subset were also somewhat higher in the present study (M = .47, SD = 0.36) than in HH (M = .22, SD = 0.30). This pattern of fewer samples but a higher switch rate actually makes sense in that HH found a negative relationship between number of samples and switch rate. We also found a negative correlation between switch rates and number of samples before stopping, r = −.25, p < .001. Another feature to notice in Figure 3 is that the switch rates for the decision policy analysis are more evenly distributed compared with the lower switch rates for the underweighting analysis. The reason, explained by the simulations in the introduction, is that divergence is more likely to occur at higher switch rates, and only scenarios with divergence were used in the analysis of decision policies.

Because of the free-sampling nature of this task, there could be potentially large differences in how many samples were drawn from each option. We calculated a "balance" metric ranging from .5 (both options were sampled the same number of times) to 1 (one option was sampled exclusively). The average balance across the 574 scenarios was .57 (SD = 0.08); participants were fairly balanced in sampling both options.

8.2.2 | Switch rate

In the subset of 574 scenarios in which all outcomes occurred at least once, we analyzed the relation between switch rate and the probability of underweighting the rare outcome using a logistic regression with by-problem and by-subject random intercepts and slopes. The maximal model failed to converge, so we dropped the correlation parameter between the by-subject intercept and slope. There was no significant relationship, B = −0.001, CI [−0.92, 0.92], p = .99. Table 2A reports the predictions of underweighting at the extremes when the switch rate is 0 and 1.

In the further subset of 258 scenarios in which the summary and round-wise policies predicted different choices, we analyzed the relation between switch rate and choosing the option predicted by the round-wise versus summary policies using a logistic regression with by-problem and by-subject random intercepts and slopes but dropped the correlation parameters between the slopes and intercepts due to convergence issues. There was no significant relationship, B = 0.51, CI [−0.35, 1.36], p = .25. Table 2B reports the predicted probabilities of choosing either option, which can be compared with Figure 3 in Hills and Hertwig (2010).

8.2.3 | RO-stopping

To test the relation between RO-stopping and tending to make choices more consistent with overweighting than underweighting, we used the subset of 574 scenarios in which the rare outcome was experienced. For Problem 1, the A^{32} outcome is the "rare" outcome, and the A^{0} and B^{3} outcomes are considered "common" outcomes (Table 1). Participants were more likely to choose the option consistent with underweighting when stopping after a common outcome (43.9%) than after a rare outcome (19.7%); Figure 4A displays the choice proportions by final outcome. We tested this effect using a logistic regression with a by-problem random intercept and slope and a by-subject random intercept. Due to the small sample of scenarios in which participants stopped sampling after the rare outcome, the maximal model failed to converge; we dropped the by-subject random slope. The effect was significant, B = −1.12, SE = 0.54, p = .04. The way to think

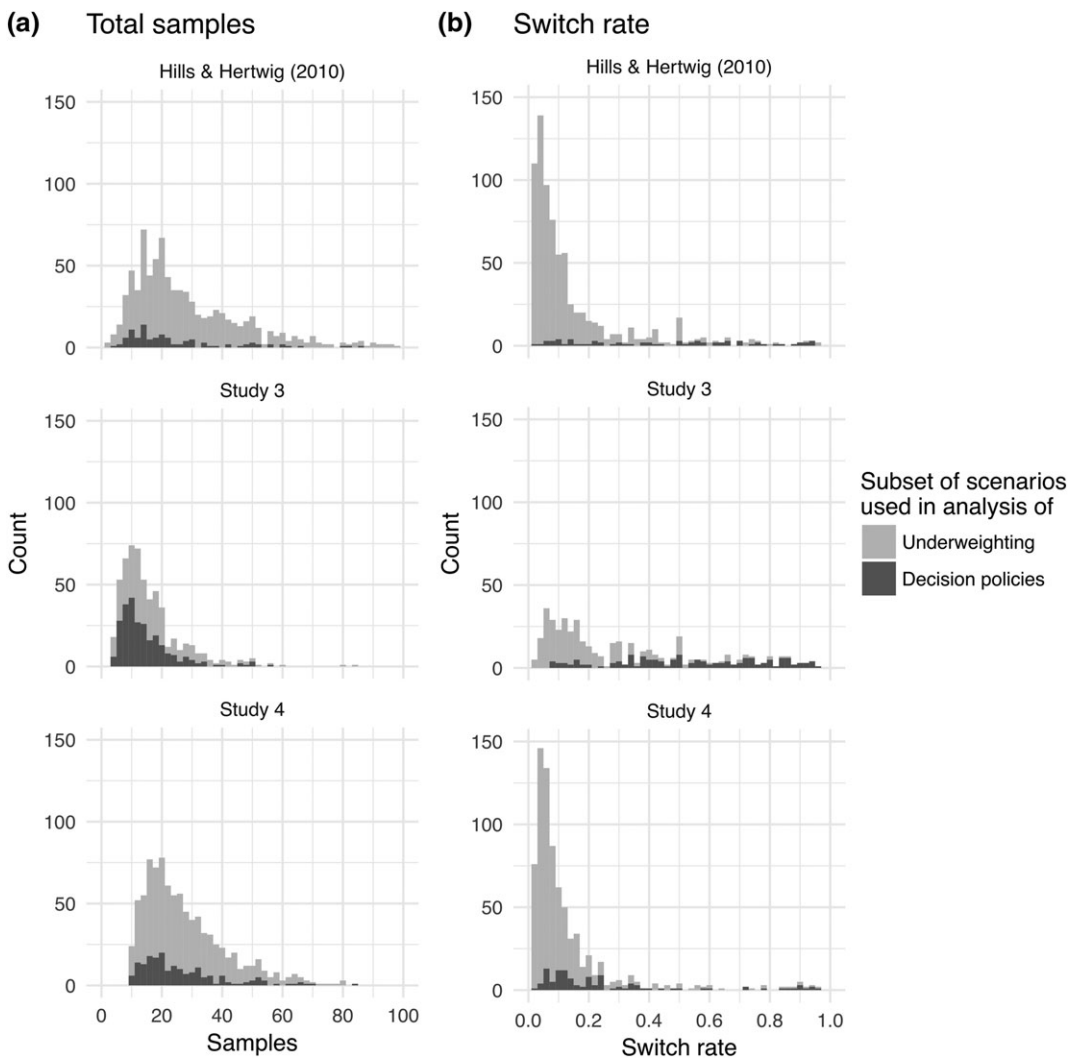

FIGURE 3 (a) Number of samples drawn in sampling phases of scenarios in which all outcomes occurred. (b) Switch rates in sampling phases of scenarios in which all outcomes occurred. The darker shade indicates the subset of scenarios in which the decision policies diverged. The HH data are at the scenario level, not aggregated at the participant level

about this finding is that participants who stopped sampling right after the rare outcome were more likely to act as if they overweighted the rare outcome compared with participants who continued to sample after experiencing the rare outcome.

To test the relation between RO-stopping and tending to make choices more consistent with the summary policy, we used the sample of 258 scenarios with divergence. Participants were more likely to make choices in line with the summary as opposed to the round-wise policy if they stopped right after the rare outcome (78.6%) than right after the common outcome (54.8%); Figure 4B displays the choice proportions by final outcome. We tested this effect using a logistic regression with a by-problem random intercept and slope, and a by-subject random intercept. Similar to the prior regression, the maximal model failed to converge; we dropped the by-subject random slope. The effect was significant, $B = 1.12$, $SE = 0.49$, $p = .02$.

8.3 | Discussion

Concerning the relationship between switch rate with underweighting and decision policies, unlike in HH, Study 3 found evidence that was, again, not significant. Furthermore, we uncovered a novel finding that participants who engage in RO-stopping tend to choose the option consistent with overweighting the rare outcome and tend to make decisions consistent with the summary policy.

9 | STUDY 4: FREE SAMPLING WITH A LAB-BASED SAMPLE

While the method in Study 3 closely resembled the free-sampling version of the DFE task used in past research (e.g., Hau et al., 2008; Hertwig et al., 2004), participants collected fewer samples and switched more frequently than in HH. It is possible that conducting the study via MTurk gave rise to this difference. In a final attempt to replicate the effects, we conducted an in-person lab-based study. While data from behavioral studies run on MTurk typically replicate most established effects (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Hauser & Schwarz, 2016; Horton, Rand, & Zeckhauser, 2011; Paolacci & Chandler, 2014), Study 4 was designed to mimic the method of the studies analyzed in HH as closely as possible.

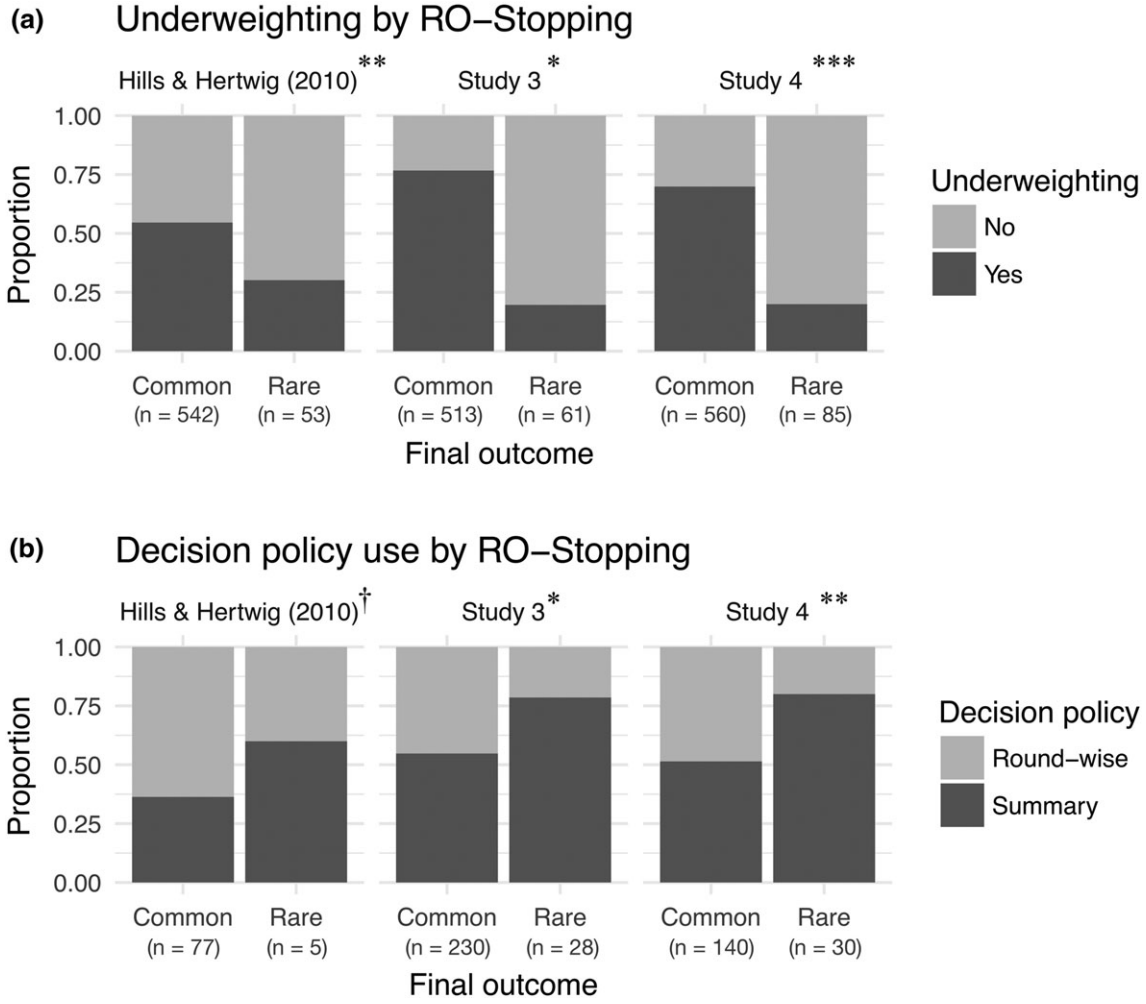

FIGURE 4 (a) Proportion of final decisions consistent with underweighting the rare outcome when the last outcome during the sampling phase was common versus rare. Optional stopping data from HH excludes Problems 12 and 13 and data from Ungemach et al. (2009) because their procedure required participants to sample a predetermined number of samples. (b) Proportion of final decisions consistent with the summary vs. round-wise decision policy when the last outcome was common versus rare. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. †Statistical test could not be run due to insufficient sample size

9.1 | Method

9.1.1 | Participants and payment

We recruited 100 Introduction to Psychology students for partial course credit. Partial data from three participants were not recorded due to a programming error. Because there was no base payment, we increased the reward for points earned; participants were paid 20 cents for each point, up to a maximum of $10. Each participant began the study with 5 points. If they finished the study with 0 or negative points, they were not paid a reward. The average payment was $3.44 (*SD* = $2.13).

9.1.2 | Design and procedure

The procedures were similar to Study 3 with the following changes. The study was conducted in a laboratory. Participants worked with all 13 decision problems in Table 1 in a randomized order and were encouraged to take breaks between decision problems if they were tired. To ensure that participants collected larger samples than in Study 3, they could only advance to the choice phase after collecting at least five samples from each option.

9.2 | Results

Initially, there were 1,266 scenarios. Eliminating scenarios in which some outcomes never occurred left 920 scenarios. For the analysis of underweighting, due to a programming error, we omitted scenarios using Problem 7, leaving 822 scenarios (65% of all scenarios).[10] For the analysis of decision policies, we analyzed the subset of the 920 in which the summary and round-wise policies diverged, resulting in 192 scenarios (15% of all scenarios).

9.2.1 | Sampling phase

Figure 3 displays the number of samples drawn and the switch rates in the sampling phases of the scenarios used in the analyses of

[10]The error resulted in the "rare outcome" of Problem 7 appearing with a probability close to .5 instead of .1 (as specified in Table 1). Because none of the outcomes in this problem were experienced as "rare," we omitted them from the analysis of underweighting rare outcomes. This error did not affect the analysis of decision policies because divergence of the decision policies can be computed regardless of whether there is a rare outcome.

underweighting (lighter gray) and decision policies (darker gray) for Study 4.

The total samples drawn in the subset of 822 scenarios used in the analysis of underweighting (M = 29.39, SD = 18.33) were fairly similar to HH (M = 31.39, SD = 25.59). The switch rates in that same subset (M = .28, SD = 0.28) were also fairly similar to HH (M = .22, SD = 0.30). Again, we found a negative correlation between switch rates and number of samples before stopping, r = −.09, p < .001. The balance metric for this subset revealed both options were sampled roughly equally (M = .58, SD = 0.08).

9.2.2 | Switch rate

We analyzed the relation between switch rate and underweighting the rare outcome in the subset of 822 scenarios, and the relation between switch rate and decision policies in the subset of 192 scenarios. Both these analyses used logistic regressions with by-problem and by-subject random intercepts and slopes. Consistent with all three prior studies, switch rate was not predictive of underweighting, B = −0.10, CI [−0.71, 0.51], p = .75 (Table 2A). Switch rate also did not predict whether participants' final decisions were more in line with the summary or round-wise policies, B = 0.29, CI [−0.70, 1.27], p = .57 (Table 2B).

9.2.3 | RO-stopping

To test the relation between RO-stopping and tending to make choices more consistent with overweighting than underweighting, we omitted scenarios in which the rare outcome never occurred, Problem 7 (due to the programming error) and Problems 12 and 13, because the difference in probability between the "rare" outcome with the next rarest outcome is very small. There were 645 scenarios after the omissions. Participants were more likely to choose the option consistent with underweighting when stopping after a common outcome (55%) than after a rare outcome (20%); see Figure 4A. We tested this effect using a logistic regression with by-subject and by-problem random intercepts and slopes. Similar to Study 3, the effect was significant, B = −1.72, SE = 0.49, p < .001.

To test the relation between RO-stopping and tending to make choices consistent with each decision policy, we used a further subset of 170 scenarios in which the summary and round-wise policies diverged. Participants were more likely to make choices consistent with the summary policy if they stopped right after the rare outcome (80%) than right after the common outcome (51%); see Figure 4B. We tested this effect using a logistic regression with a by-problem random intercept and slope, and a by-subject random intercept. The maximal model failed to converge; we dropped the by-subject random slope. Similar to Study 3, the effect was significant, B = −1.56, SE = 0.56, p = .006.

9.3 | Discussion

In Study 4, we found no relationship between switch rate and underweighting nor between switch rates on decision policies. Because this pattern of findings is consistent with Studies 1, 2, and 3, we believe that the findings are not due to methodological issues associated with the study being conducted online versus in person. In addition, similar to Study 3, we found a relationship between optional stopping with both underweighting and decision policies.

10 | GENERAL DISCUSSION

In four studies, we examined the influence of participants' switch rate and optional stopping during sampling on their final decisions. One reason for studying them together is that previous research had found that higher search rates were associated with stopping sampling earlier (Hills & Hertwig, 2012; Rakow et al., 2008), which we replicated.

Most importantly, we found that when participants stopped sampling right after experiencing a rare outcome (RO-stopping), they were more likely to choose the option consistent with overweighting the rare outcome and, similarly, were more likely to use the summary policy than the round-wise policy.

With regard to the switch rate, we tested Hills and Hertwig (2010) hypothesis that individuals who sampled with a higher switch rate tended to make decisions more in line with the round-wise than summary policy and tended to underweight the rare outcome more (see Appendix A for more details on underweighting). We ran two studies that experimentally manipulated sampling strategies between low and high switch rates and two studies using the free-sampling paradigm in which participants chose their own switch rate and when to stop sampling. All of the effects were nonsignificant. Further, Bayesian parameter estimation revealed that the most credible regression weights were centered roughly around a log odds of 0, indicating no effect (see Appendix B).

10.1 | Contributions to research on optional stopping

The current findings on optional stopping build on some important recent findings by Wulff et al. (2018). They noted that when participants get to decide how to sample and when to stop, there are recency effects; their final decisions are better predicted by the experiences near the end of sampling than near the beginning. They also simulated how stopping when there is a large discrepancy between the average outcomes of the two options can produce a recency effect (also see Coenen & Gureckis, 2016).

Our findings build upon this theory and extend it. Whereas Wulff et al. made the connection between optional stopping and recency, we have added the connection to underweighting or overweighting rare outcomes and to the use of different decision policies. The link to weighting of rare outcomes is especially relevant because much research on DFE has sought to explain the apparent underweighting of rare outcomes and identify moderating factors (Hau et al., 2008; Hertwig & Pleskac, 2010). Our analysis contributes to this endeavor. When participants stopped sampling after a common outcome, their final judgments were more likely to be consistent with underweighting. However, because rare outcomes are by definition rare, most of the time participants stop sampling after a common outcome, in which case their final judgments are less likely to overweight rare outcomes, which fits with the broader trend of underweighting.

10.2 | Methodological challenges and explanations for discrepant results for the effect of switch rate

The hypotheses put forth in HH are challenging to study because they require large amounts of data. In HH's study, only 46% of the data could be used for the analysis of underweighting, and only 6% of the data could be used for the analysis of decision policies. We used a variety of methods to increase power including collecting large samples, focusing on decision problems that produced higher rates of divergence, manipulating switch rates, and using hierarchical regression for repeated measures instead of subject-level analyses. Each of our four studies had more usable data than in HH; however, we never found significant effects.

Are there other explanations for the discrepant results compared with HH? One possibility, raised after Studies 1 and 2, was that perhaps participants must be in full control of their sampling strategies for the effect of switch rate to hold. However, Studies 3 and 4 both used a free-sampling task and still found no effect. Another possibility is that our participants' sampling behavior differed to some extent from prior studies (also see Hadar & Fox, 2009; Lejarraga et al., 2012). In Study 3, participants drew less than half the number of samples and had double the switch rate on average compared with HH. However, in Study 4, the sampling behavior was similar to HH. That both Studies 3 and 4 produced null effects, despite the differences in sampling behavior, attests to the pervasiveness of our null effects.

In sum, given the nonsignificant effects with large sample sizes, we conclude that the effects, if they exist, are small, and only occur in relatively rare circumstances.

10.3 | Theoretical advancements

Aside from the empirical findings, another main contribution of this research is the theoretical analysis and simulation presented in Figure 2, which reveals several insights. First, the simulations helped flesh out the details of HH's theory. Higher switch rates were predicted to lead participants to use the round-wise policy. At higher switch rates, the two decision policies were more likely to diverge. When the two policies diverged, the round-wise policy chose the option consistent with "underweighting" 97.6% of the time. In sum, the hypothesized mechanism for underweighting must occur through using the round-wise policy instead of the summary policy.

Second, the simulation revealed five different groups of problems which produce divergence at different rates. Most importantly, the simulations revealed that overall divergence is quite low, and when it does occur, it occurs primarily for Groups I and II. In these groups, the EVs for the two options are very similar. Practically, this means that if participants use the round-wise policy, even if they pick the option that diverges from the summary policy, this option really would not be so bad, on average. The round-wise policy is a fairly good heuristic for this set of problems.

10.4 | Possible explanations for underweighting rare outcomes in decisions from experience

Given our inability to replicate HH's explanation for underweighting, we now consider other explanations for underweighting.

10.4.1 | Optional stopping

Our optional stopping analysis revealed that when participants stopped sampling immediately after a rare outcome, they were less likely to underweight the rare outcome. However, stopping was still much more likely after a common outcome. When participants stopped sampling after a common outcome (the majority of the time), their final judgments were more likely to be consistent with underweighting. Furthermore, our analysis of optional stopping did not even consider the choices that were made when the rare option was never experienced —for example, $A^0B^3A^0B^3$ for Problem 1. In this case, participants were highly likely to choose B, which is the option consistent with underweighting the rare A^{32} outcome.[11] In sum, the majority of the time, participants stopped sampling after a common outcome or before the rare outcome occurred, and in these situations, underweighting was common.

10.4.2 | Other explanations

Other explanations have also been proposed for underweighting rare outcomes. For example, we previously discussed how drawing a small number of samples in DFE tasks can lead to sampling error, amplifying the difference between experienced outcomes and those implied by the long-run payoff distributions (Camilleri & Newell, 2013; Fox & Hadar, 2006; Hertwig & Pleskac, 2010). Some researchers have proposed that recency effects can also amplify the difference, for essentially the same reason as the small sample argument, though the evidence on this is mixed (Ashby & Rakow, 2014; Hau et al., 2008; Hertwig et al., 2004; Ungemach et al., 2009). In sum, further understanding the reasons for underweighting in DFE is a continuing task for future research.

11 | CONCLUSION

In four studies, we did not find evidence to support Hills and Hertwig's (2010) finding that participants who sampled with a high switch rate were more likely to make decisions that could be interpreted as underweighting rare outcomes. We conclude that Hills and Hertwig's intuitive and important finding that sampling strategies influence the use of decision policies and underweighting rare outcomes may not be real or may only occur in very rare circumstances. However, we found evidence for a new relation between the sampling strategy and underweighting and decision policies; participants who tended to stop sampling right after experiencing a rare outcome were less likely to underweight that rare outcome and were more likely to make decisions in line with a summary policy. These new findings provide insight into how optional stopping plays a role in decisions from experience.

ACKNOWLEDGEMENTS

This research was supported by National Science Foundation (NSF) 1430439.

The authors thank Thomas Hills for many useful conversations about this research, as well as for sharing the data and code used in

[11]Analyzing only scenarios in which the rare outcome was never experienced, participants chose the option consistent with underweighting in 97% of scenarios in Study 3, 96% of scenarios in Study 4, and 87% of scenarios in HH.

Hills and Hertwig (2010). Raw data for the studies presented here are available on the Open Science Framework: osf.io/f6qt4/.

ORCID

Kevin W. Soo http://orcid.org/0000-0002-3927-6384
Benjamin M. Rottman http://orcid.org/0000-0002-4718-3970

REFERENCES

Ashby, N. J. S., & Rakow, T. (2014). Forgetting the past: Individual differences in recency in subjective valuations from experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1153–1162. https://doi.org/10.1037/a0036352

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. https://doi.org/10.1002/bdm.443

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Camilleri, A. R., & Newell, B. R. (2013). The long and short of it: Closing the description-experience gap by taking the long-run view. *Cognition*, *126*, 54–71. https://doi.org/10.1016/j.cognition.2012.09.001

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Li, P. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Coenen, A., & Gureckis, T. M. (2016). The distorting effect of deciding to stop sampling. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2819–2824). Austin, TX: Cognitive Science Society.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review*, *114*(1), 177–187. https://doi.org/10.1037/0033-295X.114.1.177

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., … Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *The Journal of Behavioral Decision Making*, *23*, 15–47. https://doi.org/10.1002/bdm

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*(2), 159–161. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Fried, L. S., & Peterson, C. R. (1969). Information seeking: Optional versus fixed stopping. *Journal of Experimental Psychology*, *80*(3), 525–529. https://doi.org/10.1037/h0027484

Gabry, J., & Goodrich, B. (2017). Estimating generalized linear models for binary and binomial data with rstanarm. Retrived from http://mc-stan.org/rstanarm/articles/binomial.html

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, *118*(4), 523–551. https://doi.org/10.1037/a0024558

Gonzalez, C., & Dutt, V. (2012). Refuting data aggregation arguments and how the instance-based learning model stands criticism: A reply to Hills and Hertwig (2012). *Psychological Review*, *119*(4), 893–898. https://doi.org/10.1037/a0029445

Gonzalez, C., & Dutt, V. (2016). Exploration and exploitation during information search and consequential choice. *Journal of Dynamic Decision Making*, *2*(2), 1–8. https://doi.org/10.11588/jddm.2016.1.33651

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*, 591–635. https://doi.org/10.1016/S0364-0213(03)00031-4

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making*, *4*(4), 317–325.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, *23*, 48–68. https://doi.org/10.1002/bdm

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518. https://doi.org/10.1002/bdm

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience. Do our patterns of sampling foreshadow our decisions? *Psychological Science*, *21*(12), 1787–1792. https://doi.org/10.1177/0956797610387443

Hills, T. T., & Hertwig, R. (2012). Two distinct exploratory behaviors in decisions from experience: Comment on Gonzalez and Dutt (2011). *Psychological Review*, *119*(4), 888–892. https://doi.org/10.1037/a0028004

Hills, T. T., & Hertwig, R. (2017). Corrigendum: Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science*, *28*(9), 1364–1366. https://doi.org/10.1177/0956797617717946

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*, *20*(5), 1023–1031. https://doi.org/10.3758/s13423-013-0422-3

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*, 399–425. https://doi.org/10.1007/s10683-011-9273-9

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312. https://doi.org/10.1177/1745691611406925

Kruschke, J. K. (2015). *Doing bayesian data analysis: A tutorial with R, JAGS, and Stan*. New York: Academic Press.

Kruschke, J. K., & Liddell, T. M. (2016). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and planning from a Bayesian perspective. *Psychonomic Bulletin & Review*, 1–53. https://doi.org/10.2139/ssrn.2606016

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 856–876. https://doi.org/10.1037/0278-7393.30.4.856

Lejarraga, T., Dutt, V., & Gonzalez, C. (2010). Instance-based learning: A general model of repeated binary choice. *The Journal of Behavioral Decision Making*, *25*(2), 143–153. https://doi.org/10.1002/bdm

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*(3), 334–342. https://doi.org/10.1016/j.cognition.2012.06.002

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122. https://doi.org/10.1037/a0032108

Markant, D. B., Pleskac, T. J., Diederich, A., Pachur, T., & Hertwig, R. (2015). Modeling choice and search in decisions from experience: A sequential sampling approach. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1512–1517). Austin, TX: Cognitive Science Society.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215. https://doi.org/10.1037/dec0000033

Noguchi, T., & Hills, T. T. (2015). Experience-based decisions favor riskier alternatives in large sets. *Journal of Behavioral Decision Making*, *29*(5), 489–498. https://doi.org/10.1002/bdm.1893

Noguchi, T., & Hills, T. T. (2016). Description-experience gap in choice deferral. *Decision*, *3*(1), 54–61. https://doi.org/10.1037/dec0000044

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019. https://doi.org/10.1037/xge0000013

Plonsky, O., & Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. *Cognitive Psychology*, *93*, 18–43. https://doi.org/10.1016/j.cogpsych.2017.01.002

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, *106*, 168–179. https://doi.org/10.1016/j.obhdp.2008.02.001

Rottman, B. M. (2016). Searching for the best cause: Roles of mechanism beliefs, autocorrelation, and exploitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1233–1256.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Stan Development Team (2017). RStanArm: Bayesian applied regression modeling via Stan. Retrived from http://mc-stan.org

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*, 219–225.

Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/Bf00122574

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*(4), 473–479. https://doi.org/10.1111/j.1467-9280.2009.02319.x

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430–445. https://doi.org/10.1037/0033-295X.111.2.430

Wegier, P., Bianchi, L. J., & Spaniol, J. (2015, July). Patterns of search in experiential sampling: Investigating piecewise search. Poster presented at the meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Ottawa, Canada.

Wegier, P., & Spaniol, J. (2013, June). To switch or not to switch: Search strategies in experience-based judgments of proportion. Poster presented at the meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Calgary, Canada.

Wegier, P., & Spaniol, J. (2014a, July). Patterns of information search in experience-based choice. Poster presented at the meeting of the Cognitive Science Society, Quebec City, Canada.

Wegier, P., & Spaniol, J. (2014b, November). Better and faster information search: The robustness of piecewise sampling. Poster presented at the meeting of the Society for Judgment and Decision Making, Long Beach, CA.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). How short- and long-run aspirations impact search and choice in decisions from experience. *Cognition*, *144*, 29–37. https://doi.org/10.1016/j.cognition.2015.07.006

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

How to cite this article: Soo KW, Rottman BM. Switch rates do not influence weighting of rare events in decisions from experience, but optional stopping does. *J Behav Dec Making*. 2018;1–18. https://doi.org/10.1002/bdm.2080

APPENDIX A

ALTERNATE ANALYSIS OF UNDERWEIGHTING

In Hills and Hertwig's original paper (2010), when analyzing the relation between switch rate and underweighting, they dropped scenarios in which participants never experienced the rare outcome. For example, for Problem 1, they dropped a scenario if the outcome of 32 was never experienced. There were three reasons for dropping these cases: First, when the extreme outcome was not experienced, the scenario appears deterministic; A always produces 0, and B always produces 3. In such a case, it would be bizarre for any participant to choose A. Second, there is no divergence in such cases.

Third, it would be strange to label one choice as "underweighting the rare outcome" if the rare outcome never occurred. For example, suppose that a learner experiences $A^0A^0A^0B^3B^3B^3$ and chooses B. In Problem 1, A has an outcome of 32 with probability .1, so choosing B is categorized as being consistent with underweighting the rare A^{32} outcome even though it was never experienced. However, if instead the unexperienced rare outcome is A^{20} instead of A^{32} such that the EV_A after a certain number of samples is less than 3, then choosing B would be viewed as the correct decision, not underweighting the rare outcome. This example shows that labeling one choice as

"underweighting" when an option is never experienced critically depends on the decision problem.

Pleskac and Hertwig (2014) have argued that in many real-world situations, there is a "risk–reward" tradeoff implying that if one of the two options in the $A^0A^0A^0B^3B^3B^3$ sequence has a rare positive reward, it would be A, which justifies labeling choosing B as consistent with underweighting. Indeed, researchers have tended to create decision problems in which the risk–reward heuristic holds.[12] However, this risk–reward heuristic is not a guarantee, and since researchers have not attempted to systematically sample the decision problems, we do not feel that it alone justifies treating the choice of B as consistent with underweighting, though other researchers may disagree (for additional discussion, see Wulff, Hills, & Hertwig, 2015).

In HH's correction (Hills & Hertwig, 2017), they performed the underweighting analysis with all the problems even if the rare outcome was not experienced, and found a significant effect such that underweighting was more likely with a higher switch rate. Even though we have qualms with this analysis, we conducted this analysis for Studies 2–4. For Study 1, this analysis was not necessary because all participants experienced the rare outcome. None of these studies found the effect reported in HH's correction.

In Study 2, there was a negative relationship between switch rate and underweighting (B = −0.54, 95% CI [−1.04, −0.04], p = .04), the opposite of HH's correction. In Study 3, there was no relationship between switch rate and underweighting (B = −0.43, 95% CI [−0.11, 0.96], p = .12). In Study 4, there was also no relationship between switch rate and underweighting (B = 0.23, 95% CI [−0.41, 0.86], p = .48).

APPENDIX B
BAYESIAN PARAMETER ESTIMATION

B.1 | Motivation

According to standard null-hypothesis significance testing (NHST), our findings concerning the influence of switch rate on underweighting and decision policy usage can only reject the null hypothesis or fail to reject the null hypothesis (for a discussion of the other weaknesses of NHST and proposed alternative approaches, see Cumming, 2014; Kruschke & Liddell, 2016; Rouder, Speckman, Sun, Morey, & Iverson, 2009). To test if there is evidence in favor of the null and to quantify the most credible parameter estimates, we conducted Bayesian parameter estimation (see Kruschke, 2011; 2015). This analysis produces a posterior probability distribution over possible parameter values, and the 95% highest density interval (HDI) contains the 95% most credible parameter values.

B.2 | Priors

Since our analyses were logistic regressions with coefficients on the log-odds scale (zero indicates no effect), we used a weakly informative prior—a Cauchy distribution with center = 0 and scale = 2.5 (Gelman, Jakulin, Pittau, & Su, 2008). The prior on the group specific random slopes and intercepts used a zero-mean random multivariate Gaussian distribution with a covariance matrix estimated from the data (Gabry & Goodrich, 2017). Inference was performed with the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015) and rstanarm (Carpenter et al., 2017; Stan Development Team, 2017) for R.

B.3 | Inference and interpretation

Figure B1 plots the posterior distribution of the regression weight B for both analyses across all studies. The most credible values have the highest posterior weights, and the 95% HDI of each parameter is denoted with the horizontal line. One possible inference strategy is to simply observe the most credible posterior parameter weights. Another, similar to NHST, is to assess whether the 95% HDIs ever exclude the null-hypothesis parameter zero.

Another inference possibility is to determine whether there is evidence in favor of the null hypothesis. There are two steps to doing this. First, one must establish a range of parameter values around the null value of zero that we deem to be practically equivalent to zero, called a *region of practical equivalence* or ROPE. Figure B1 displays multiple ROPEs, allowing the reader to choose their own ROPE. The ROPEs are defined on the log-odds scale; B = 0 means there is no change in the probability of choosing a particular option as the switch rate increases from 0 to 1. We also transformed the log odds to implied differences in probabilities since it can be difficult to interpret log odds. We centered the probabilities around .5, because most of the judgments had means relatively close to .5. A log odds of 0.4 corresponds to a probability difference of .1; this means that the probability of underweighting is roughly .45 when the switch rate is 0, and is roughly .55 when the switch rate is 1. A log odds of 0.81 corresponds to a probability difference of .2; this means that the probability of underweighting is roughly .4 when the switch rate is 0, and is roughly .6 when the switch rate is 1.

Second, one must look at where the HDI falls in relation to the ROPEs. If the HDI falls entirely within a ROPE, one accepts the null value at that particular level of equivalence (i.e., if the HDI falls within the ROPE of ±.10, we accept a conservative null value). If the HDI falls entirely outside a ROPE, we can reject the null value at that level of equivalence. In many cases, there is partial overlap between the HDI and a ROPE, in which case the analysis does not provide strong evidence in favor of the null or the alternative hypothesis.

B.4 | Results

B.4.1 | Analysis of underweighting

The HDI for HH's analysis does not fall entirely inside the narrowest ROPE, but it does fall entirely within the −.20 to +.20 ROPE. The most credible parameter values are relatively close to 0. This makes sense in that the reanalysis by HH was not significant.

In Studies 1 and 2, the HDI of B falls within the ±.20 ROPE. This suggests that we can accept the null hypothesis if the change in probability is no greater than .2 when comparing underweighting between

[12] We thank Thomas Hills for pointing this out.

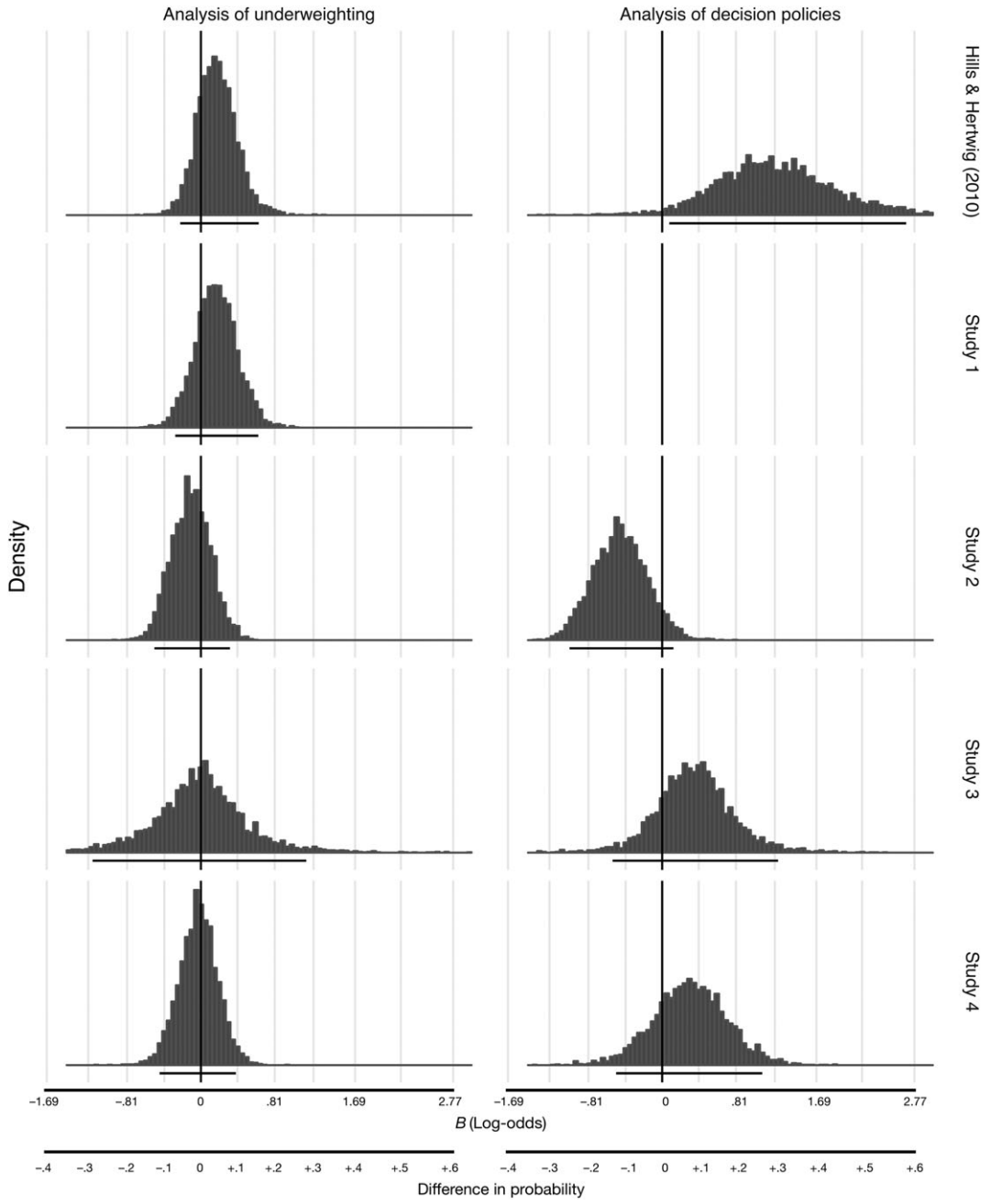

FIGURE B1 Posterior distributions of parameter values for analyses reported in Tables 2A and 2B. The 95% highest density interval is indicated by the solid line beneath each distribution. The vertical lines represent the region of practical equivalence

the low and high extremes of switch rates. In Study 3, the HDI is much more spread out and falls within the ±.30 ROPE. In Study 4, the HDI falls almost completely within the ±.10 ROPE. Taken together, the most credible values for the underweighting analysis are relatively close to zero.

B.4.2 | Analysis of decision policies

The 95% HDI for HH's study is entirely above 0, which corresponds to the standard interpretation of a "significant" effect; the most credible values are roughly centered around $B = 1.24$, which corresponds to a difference of probability of .3.

The most credible parameter values are near zero; however, none of the studies fall within the ±.20 ROPE. In summary, this analysis does not provide strong evidence in favor of the null hypothesis or the alternative hypothesis.