

Biased Preferences Through Exploitation: How Initial Biases Are Consolidated in Reward-Rich Environments

Chris Harris
Utrecht University

Klaus Fiedler
University of Heidelberg

Hans Marien and Ruud Custers
Utrecht University

In the current article, we test the prediction that an initial bias favoring 1 of 2 equally rewarding options—either based on a genuine contingency or a pseudocontingency in a small sample of initial observations—can survive over an extended period of further sampling from both options, when the reward structure fosters exploitation. Specifically, we argue and demonstrate that in reward-rich environments where two options predominantly—but equally frequently—yield positive outcomes, the initial bias should be upheld because exploitation of the allegedly superior option reinforces the biased preference. In contrast, in reward-impoverted environments, where both options yield predominantly negative outcomes, initial biases can be expected to be eradicated through exploration, which increases the chance of recognizing the equality of the initially nonpreferred option. In 3 experiments, initial evidence in a guided-sampling phase was set up for participants to perceive an actual contingency (Experiment 1) or infer a pseudocontingency (Experiment 2a and b) that made 1 option look more rewarding. In a subsequent free-sampling phase this led to a sustained bias toward this option when the environment contained mostly positive but not when it contained mostly negative outcomes. We argue that biased sampling in reward-rich environments could be responsible for false beliefs about the outcomes of behavioral options, and as such could be relevant to a broad range of topics including social interactions or health contexts.

Keywords: exploration-exploitation trade-off, action-outcome learning, initial evidence


Supplemental materials: <http://dx.doi.org/10.1037/xge0000754.supp>

A fundamental trade-off all agentic organisms face is the choice between information search and reward maximization. Only through explorative information search do they acquire more knowledge about an environment and learn, for example, about areas to avoid or where the best food sources can be found. Once a certain option is regarded as superior, however, one can then commit to this option, exploiting the payoffs this option has to

offer, and maximize the rewards in the here and now (Mehlhorn et al., 2015). For example, if you move to a new town you most likely engage in exploration where you try out many different restaurants, forming an initial belief regarding the quality of these establishments. Yet, even in the largest of cities we would eventually return to a restaurant we liked and start visiting it more often than alternatives. Balancing these two adaptive functions, exploration and exploitation, is relevant in all repeated choices and is essential for maximizing positive outcomes (Cohen, McClure, & Yu, 2007; Hills, Todd, Lazer, Redish, & Couzin, 2015).

However, the temptation to exploit a seemingly superior option may prevent people from obtaining a good balance between exploration and exploitation. Instead, premature exploitation increases the risk that they fail to engage in sufficiently long exploration to figure out the alternative that is more attractive in the long run. In the research reported in the current article, we emphasize the power early experience in a sequential contingency assessment task can have in inducing a sustained and biased exploitation strategy that is very hard to attenuate even after an extended series of observations. Such a dysfunctional primacy effect can prevent people from gathering valuable information about other options and create and maintain an

This article was published Online First April 23, 2020.

 Chris Harris, Department of Psychology, Utrecht University; Klaus Fiedler, Department of Psychology, University of Heidelberg; Hans Marien and Ruud Custers, Department of Psychology, Utrecht University.

Insightful comments by Toby Pilditch and Henk Aarts are gratefully acknowledged. Parts of this article were presented in 2019 at the Subjective Probability, Utility, and Decision-Making Conference (SPUDM) in Amsterdam and the Social Cognition Network (ESCON) Transfer of Knowledge Conference in Bordeaux. All data and material can be found on the following repository: <https://osf.io/j3ve4/>.

Correspondence concerning this article should be addressed to Chris Harris or Ruud Custers, Department of Psychology, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, the Netherlands. E-mail: c.a.harris@uu.nl or r.custers@uu.nl

illusion of contingency in favor of the exploited alternative, despite the fact that an extended series of observations does not support such a contingency.

Specifically, we demonstrate that choice behavior as well as relative contingency and conditional probability estimates that are informed by an extended series of observations depend on a distinct interaction between two influences: the primacy effect of an early contingency inference and the subsequent influence of reward-rich versus a reward-impooverished environment on the tendency to engage in premature exploitation. In a reward-rich environment, in which the absolute base rate of positive outcomes is high for both alternatives, the temptation to exploit the seemingly better alternative is reinforced again and again, creating little reason to switch and learn about the other alternative. This should render the premature preference highly stable and defer any tendency to correct for the primacy effect. In contrast, when the absolute rate of positive outcomes is low in a reward-impooverished environment, the motive to stick to the alternative favored by the early contingency inference is rather low, rendering exploration and correction for the premature early inference more likely.

Origins and Maintenance of Biased Contingencies

To understand the theoretical rationale underlying the depicted two-stage process, let us first discuss the cognitive origins of contingency inferences based on a few initial observations, before we turn to the maintenance of early contingencies in reward-rich and reward-impooverished environments.

Beliefs Originating in Contingencies Inferred From Small Samples

Barring outside influences such as information from peers or other sources, one would have to form an initial belief on an available sample of first evidence. Such a belief will likely be based on a small initial sample. However, small samples have the statistical property of inflating correlations (Kareev, 1995, 2000; Kareev, Lieberman, & Lev, 1997). This is crucial given the well-established finding that the order of evidence has a strong influence on the inferences people draw from sequential information sampling. Especially the first few trials often influence people particularly strongly (primacy effects; Anderson, 1965; Asch, 1946; Dennis & Ahn, 2001; Jones, Goethals, Kennington, & Severance, 1972) in their learning as well as their behavior (Decoster & Claypool, 2004; Pilditch & Custers, 2018; Pilditch, Madson, & Custers, 2020; Staudinger & Büchel, 2013). Random influences or short-term fluctuations in the environment could thus induce biased contingencies inferred from small initial samples.

Even when small samples do not provide the information required for genuine contingency inferences, pseudocontingencies can be inferred heuristically from skewed attribute base rates (Fiedler, 2010; Fiedler, Freytag, & Meiser, 2009; Meiser, Rummel, & Fleig, 2018). In most environments some events are more prevalent than others. And often one action is performed more frequently than the alternative. In other words, base rates are skewed as a decision maker executes one action more often than another and one event is more likely to occur. Such a double-skewed situation gives rise to pseudocontingency inferences. The more (less) frequent level of one variable appears to co-occur with

the more (less) frequent level of the other variable (Meiser & Hewstone, 2006). Thus, pseudocontingency inferences confuse contingencies at different aggregation levels. If in an ecological setting one option is more prevalent than another, and one outcome is more prevalent than another, this alignment of base rates seems to suggest a causal influence of the prevalent cue on the prevalent outcome. Such an inference is of course unwarranted, logically, but it can be shown that the pseudocontingency heuristic predicts the true contingency most of the time (cf. Fiedler, Kutzner, & Vogel, 2013). Nevertheless, heuristic inferences of the pseudocontingency type constitute a frequent source of quick and premature contingency estimates. In summary then, biased or premature contingency inferences may be induced quickly, regardless of whether the first few observations exhibit a genuine contingency or nothing but aligned base rates.

Belief-Maintenance Depends on Reward Experience

But while such primacy effects can readily explain the emergence of initial erroneous beliefs, one might expect such biases to wear off or to be corrected during repeated sampling. Such attenuation is, however, not so self-evident. Numerous accounts speak to the almost trivial assumption that hedonic experiences influence subsequent information search and choices (Denrell, 2005; Denrell & Le Mens, 2012; Fazio, Eiser, & Shook, 2004; Higgins, 1997; Lave & March, 1993; Skinner, 1948; Thorndike, 1927).¹ Consistent with the game-theoretical maxim “win-stay-lose-shift,” individuals tend to exploit and stick to a preferred alternative if it is often met with success, but switch to another alternative if outcomes are often negative. In other words, an initially preferred alternative is more likely to be retained when hedonically positive outcomes are frequent, whereas an infrequently rewarded alternative is more likely to be given up.

For good reason people typically do not fully maximize even under exploitation schemes as the binary description so far might suggest. Instead, a typical finding is probability matching. That is, people keep exploring to some degree by matching their choice probabilities to the outcome probabilities of interest (Vulkan, 2000). However, even when decision makers engage in probability matching, sacrificing part of their decision trials for exploration, the exploitation behavior would typically still uphold the skewness of the previous distribution. The favorable option would still be sampled more often than the alternative. Such hedonic sampling tendencies are crucial to understanding the persistence of biased initial contingencies, regardless of the underlying cognitive algorithm (genuine contingency inference or pseudocontingencies).

Reward-Rich and Reward-Impooverished Environments

An appropriate experimental manipulation of hedonic influences on belief maintenance is to contrast reward-rich and reward-impooverished environments. In a reward-rich environment, in which outcomes are usually positive and less often negative, an initial contingency inference is more likely to be maintained and

¹ To be sure, hedonic value is not the only determinant of information sampling; another determinant is the epistemic value or diagnosticity (Prager, Krueger, & Fiedler, 2018). In the present investigation, though, the focus is on the hedonic influence of reward.

an erroneous contingency is less likely to be corrected than in a reward-impooverished environment with outcomes being more often negative and less often positive. In a reward-impooverished environment, it is equally conceivable that one would form biased preferences off an initial sample. Instead of favoring one option following a few positive experiences, a decision maker might dislike an option following a few negative experiences and shift toward the other option. However, as further sampling would reveal that the alternative option also leads to negative experiences, we argue that a decision maker would not stick to this option. No option would appear to be a clear favorite warranting exploitation. Instead, a decision maker would have to extend their exploration resulting in oscillation between the choice alternatives. This, in turn, would undo the skews of the options chosen and thereby undo the prerequisite for pseudocontingency inferences. Without frequent hedonically positive outcomes, exploitation and the maintenance of initial biases is less likely. In other words, we claim that in reward-rich environments participants' skewed sampling behavior upholds initial biases while in reward-impooverished environments participants' sampling behavior undoes the skews and initial biases are mitigated.

In any case, the formation and maintenance of biased beliefs not only hinges on the individual's preferred strategy for regulating the trade-off between exploration and exploitation. It also depends on the environment decision makers find themselves in. While a reward-rich environment would encourage exploitative behavior, a reward-impooverished environment would not, crucially shifting the trade-off between exploration and exploitation. This process is important to understand as it can explain how under certain conditions, but not others, people are likely to uphold initial biases even though they continue gathering more information.

Backup From Learning Models

The assumptions about the moderating impact of the probability of rewarding outcomes on the maintenance of initial contingency inferences receive convergent support from several learning models, which all predict lesser belief change with increasing sampling from the initially preferred alternative.

BIAS (Brunswikian Induction Algorithm for Social Inferences; Fiedler, 1996) is a connectionist model that represents concepts and stimuli in distributive format, as patterns or vectors of sub-symbolic elements. Within this framework, a series of the observations of positive or negative outcomes of decision Options A and B would be represented as a matrix, the columns of which represent the learning trials. Each column contains vector segments for the option (i.e., a noisy copy of the pattern defining A vs. B) concatenated with a vector segment for the outcome (i.e., a noisy copy of the patterns denoting positive vs. negative outcomes), maybe along with other vector segments (for context, time, etc.). Thus, at the end of the learning process, the matrix contains noisy representations of the entire stimulus input. An evaluative judgment (of the degree to which outcomes of a particular option have been positive) is simulated by prompting the matrix with the ideal pattern of one option, say, A. The BIAS algorithm then compares all matrix columns in the option segment with that prompt and multiplies each column vector with the cross product that quantifies the match with the prompt. In this way, those vectors in the memory representation that are relevant to the prompted option

receive a higher weight than irrelevant matrix columns. The weighted row means across all matrix columns in the outcome segment is then compared to the ideal type of positive (vs. negative) outcomes. The higher the correlation between the weighted row means (in the outcome segment) and the ideal type, the more positive is the simulated judgment. It is easy to show that the size of this correlation increases with the number of columns that speak to the focal option. In other words, the prevailing valence is accentuated when an exploitation strategy increases the number of observations about the focal alternative.

Another learning model that would make similar predictions is Minerva-DM (MDM; Dougherty, Gettys, & Ogden, 1999). It is quite similar to BIAS, with differences between the two models being mainly in the interpretation of the underlying processes. Importantly though, just as BIAS, MDM probes a (noisy) memory structure and aggregates the matches. And so, once again, an exploitation strategy would increase the number of matrix columns reflecting the predominant trend and thus result in the prevailing valence being accentuated.













It is important to highlight the contrast between models such as BIAS and MDM on the one hand, and models based on updating of a variable of interest such as expected utility. The latter include many reinforcement learning models which typically do not retain sensitivity for the underlying base rates as they instead update singular value. In other words, the models typically do not retain any information on whether an estimate of expected utility is based on a sample of 10 or 100 trials and exploitation would not result in an advantage for either of two equally good options.

The Present Studies

In the remainder of this article, we test our claim that erroneous contingency inferences are formed and maintained in reward-rich but attenuated in reward-impooverished environments with simulations as well as several experiments. We do so by using a two-armed bandit task, in which participants repeatedly chose between two bags, A and B, for a total of 100 trials. Choosing either bag resulted in either of two outcomes, a blue or a yellow ball being grabbed. One of these colors would result in participants winning points, while the other color would result in participants losing points. Critically, the distribution of blue and yellow balls was identical in both bags. There were two conditions: A reward-rich condition, in which each bag contained 75 winning and 25 losing balls, and a reward-impooverished condition, in which each bag contained 75 losing and 25 winning balls.

The 100 trials can be divided into two phases, an induction phase and a free sampling phase. In the induction phase, participants were forced to select the two bags a certain number of times. Additionally, we controlled the outcomes, such that participants encountered a specific distribution of initial evidence. In Experiment 1, this evidence consisted of four trials that suggested a perfect genuine contingency between bags and ball color. In Experiments 2a and b, participants would encounter a distribution of 16 trials that contained a zero contingency between bags and ball color, but skewed base rates yielded a distinct pseudocontingency. We used these same two distribu-

Table 1
Distributions of Initial Evidence

Condition		Experiment 1		Experiment 2	
Reward-rich  	Frequently shown bag	 Wins 3	 Losses 0	 Wins 9	 Losses 3
	Infrequently shown bag	0	1	3	1
Reward-impooverished  	Frequently shown bag	 Wins 0	 Losses 3	 Wins 3	 Losses 9
	Infrequently shown bag	1	0	1	3

Note. The distributions used for the initial evidence in the induction phase. Also depicted are the images used to denote the choices and outcomes in the studies. See the online article for the color version of this table.

tions of initial evidence (see Table 1) for the simulations of three different learning models in a simulation study.

Simulations: Learning Model Simulations

We compared the learning model discussed above, BIAS, to a basic Bayesian learning model, and an updating-based reinforcement learning model. The Bayesian model consists of two independent arms for the two options from which the better alternative is estimated. While slightly more complex than a Bayesian model with two dependent options, we were interested in this Bayesian framework, as these are oftentimes considered normative accounts of how information updating should take place in decisions under uncertainty. For a reinforcement learning model, we simulated here a Rescorla-Wagner learning model (Rescorla & Wagner, 1972) for which an expected value function is updated according to the prediction error between predicted and experienced outcome. Importantly, the Rescorla-

Wagner model does so with the same weight after the first trial as it does after the 100th trial, rendering this model insensitive to the base rates sampled. While this list of learning models is certainly not exhaustive, it should serve as demonstration of the influence of initial evidence on subsequent sampling.

Method

All simulations were run using R (R Core Team, 2018) and, along with the analyses for the later studies, can be found on an Open Science Framework repository (Harris, 2020). All simulations were run $N = 10,000$ times at which point patterns were highly stable.

The simulations closely represented the task described above which participants encountered in the later studies. We feed each of the simulations initial evidence that consists of four or 16 trials also later utilized in Experiments 1 and 2 (summarized in Table 1). Then, for the remaining trials, the learning models determine the

next choice according to their respective algorithms. One assumption we therefore make here is that on each trial the models choose the option considered better. Importantly, we assume that just as in Experiments 2a and 2b, the models do not know the outcome scheme yet and prompt for the frequent option. Then, during the free sampling phase they prompt for winning as by that time they would have been informed about the outcome scheme.

For BIAS, the memory matrix consisted of one column for each trial encountered (therefore $n_{\text{col}} = 100$). On any given trial each previous instance (column) was prompted with the idealized (non-noisy) vectors for Options A and B. Mathematically, we calculated the Spearman correlation between the idealized (non-noisy) vector and the (noisy) memory instance. The resulting vector that numerically described the similarity between the ideal and the memory instance then constituted the weight that was applied to the ensuing matching between the prompt for winning and the vectors in the matrix. That is, at this time the Spearman correlation between the vectors for winning and the memory instance was calculated. The product of these two correlations was then used as the final match between a given memory instance and winning with a respective option. The option with the overall higher match was chosen for the next trial. The noise rate was set to $L = .33$ such that each cell in the column-vector would be distorted, that is ones becoming zeros and vice versa, with probability L .

For the Bayesian model, on each trial the probability density function, denoted by a beta distribution with parameters α and β that was updated on each trial as new evidence was experienced, was calculated for both arms of the bandit. An important distinction to make here, is that we calculated separate probability density functions as the two arms of the bandit were independent from one another. In other words, because the two Options A and B could in theory have probabilities independent of one another, the simulation kept track of one probability density function for Option A that described the estimated probability of winning with this option. And it kept track of an independent, second probability density function for Option B. For both of these probability density functions, the x -axis represents all possible probabilities for winning with this arm of the bandit ($0 \leq x \leq 1$) and the y -axis the respective likelihood given the data. These two independent probability density functions necessitated some form of comparison in order to determine the better option to choose on any given trial. We opted for a comparison of sampled means. So, for each arm of the bandit, three “ x ” values were sampled with probability “ y ” and their mean was then regarded as the estimated probability for winning with this arm. The arm with the higher estimated probability was chosen on this trial. The arm with the more extreme and higher maximum was therefore more likely to be considered the better alternative but the process was somewhat noisy due to the sampling of these estimates. While this model offers only a crude estimation of people’s choices over time, it offers a reasonable normative estimate of what choice a Bayesian learner might make on any given trial given the evidence encountered thus far. The number of x values sampled on each trial is arbitrary but can be thought of as parameter which serves as proxy for base rate sensitivity or certainty: The more samples are drawn from each probability density function, the more likely the resulting mean estimate approximates the x value for which the function is at its maximum, while a single draw is more likely to fluctuate across the entire range of x values.

Finally, for the Rescorla-Wagner model, we used the following, arbitrarily chosen parameters $\alpha = .25$, $\beta = 1$ with α describing the learning rate and β describing the rate of exploration. We ran the simulations for a large range of parameter combinations in order to confirm the robustness of our results (not shown here). In summary, the parameters influence just how quickly the model stabilizes around chance level which in any case takes place over a very short time period across a large range of parameter combinations. For the Rescorla-Wagner model with the initiating parameter $V_{s, 0} = 0.5$, we used the following updating Equation (1):

$$V_{s,t} = V_{s,t-1} + \alpha(r_{t-1} - V_{s,t-1}) \quad (1)$$

Along with the following observation Equation (2):

$$p(s) = \frac{\exp(\hat{a} \cdot V_s)}{\sum_i \exp(\hat{a} \cdot V_i)} \quad (2)$$

For each of these three models we ran simulations using the parameters participants would encounter in Experiments 1 and 2. We always compare a reward-rich to a reward-impooverished condition in which the probability of a positive outcome with either option is .75 or .25, respectively. Each model is fed a distribution of initial evidence before then running freely over the remaining trials. The first distribution we exposed the models to, was a perfect contingency over four trials: One option would win three times and the other option would lose once. This distribution was later applied again in Experiment 1. The other distribution was a distribution with a contingency of zero but in which one option had to be chosen more often than the alternative. It consisted of nine wins and three losses with one option and three wins and one loss with the other option in the reward-rich condition and the reversal (nine losses and three wins with one option and three losses and one win with the alternative) in the reward-impooverished condition (see also Table 1). This distribution was later applied again in Experiment 2. The order of trials for the initial evidence was randomized.

Results and Discussion

As can be seen in Figure 1, simulations using BIAS resulted in the pattern we predict. That is, the initial evidence led to strong biases and those biases were maintained during continued sampling in the reward-rich but not so in the reward-impooverished condition. In the Bayesian learning model, simulations of the reward-rich condition resulted in the predicted pattern, while simulations of the reward-impooverished condition also resulted in the maintenance (albeit far weaker) of biases. For the Bayesian model it is important, however, to point out that the simulations here do not include any error term as the other models do via noise or learning rates. The Rescorla-Wagner model, finally, which is insensitive to the base rates of the sampled distributions, did not result in the pattern predicted, but instead quickly converged to chance level. How quickly this convergence takes place would be dependent on the model specifics. Nonetheless, as the Rescorla-Wagner keeps track only of the expected rewards for the respective options and does not take into account how often a given option has been sampled, the model will become indifferent when the options are identical as is the case here. Note, that in our case this correctly describes the two (identical) options the model can

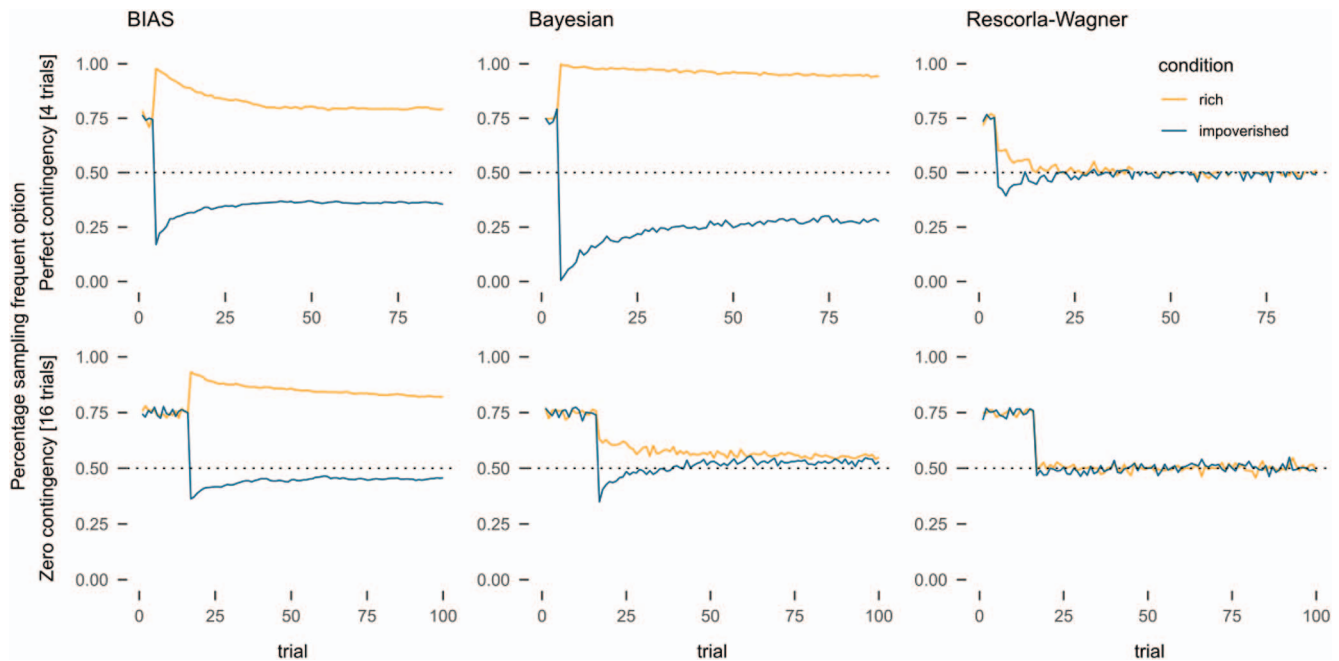


Figure 1. Probability of choosing the frequent option over the alternative in a reward-rich (green) and a reward-impooverished (orange) condition for the three learning models BIAS, Bayesian learning, and Rescorla-Wagner (left to right). The distribution of initial evidence is identical to that used in Experiment 1 in the first row and identical to that of Experiment 2 in the second row. BIAS = Brunswikian Induction Algorithm for Social Inferences. See the online article for the color version of this figure.

choose from. The Rescorla-Wagner most certainly is not “wrong,” but simply describes a different mechanism in how human learning might take place while the pattern we predict in this article relies on base rate sensitivity.

Interestingly, for the Bayesian model we can control the extent to which it is sensitive to base rates as described above. And indeed, as we increase the number of draws made to estimate the probability of winning on a given trial, the more distinct the pattern (maintained bias in the reward-rich, attenuation in the reward-impooverished condition) becomes (not shown here).

While it is true for each of these models that assumptions were made that could very well differ from how humans make choices in real-life situations, the general pattern of the learning models sensitive to base rates is striking. BIAS and the Bayesian learning model predict a bias toward one option over the alternative in the reward-rich condition while they predict such a bias to be strongly reduced if not even absent in the reward-impooverished condition. In what follows, we empirically test these same predictions.

Experiment 1: Perfect Contingency Initial Evidence

Experiment 1 set out to validate the experimental setup and the main assumption underlying this article. Namely, that a skewed distribution in the initial evidence, where one option is more frequent than the alternative and one outcome is more frequent than the alternative, would lead to biased sampling and biased preferences in a reward-rich but not in a reward-impooverished environment, given that both options are equally rewarding. The sample size was estimated to be around 100, based on power

analyses using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). These calculations are based on a 5% alpha-level, 80% statistical power, and effect sizes between $\eta_p^2 = .081$ and $\eta_p^2 = .270$ as reported by Meiser, Rummel, and Fleig (2018).² In this as well as the later experiments, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the experiments (Simmons, Nelson, & Simonsohn, 2012).

Method

Participants for this study were recruited via the online crowdsourcing platform Prolific Academic (<https://prolific.co/>) and the study was run in English on Soscisurvey (Leiner, 2014). One-hundred participants ($N_{\text{female}} = 49$) with an average age of 30 years ($SD = 6.52$) participated for a financial reward of £1.15 plus additional earnings (mean £1.40, max £1.60) based on performance. All participants indicated to be fluent in English and had an approval rating of 95 (out of 100) or higher on the platform. Ninety-one percent of participants had an educational degree of College/A levels or higher. The research line reported in this article was conducted according to the guidelines of the Ethics Review Board of the Faculty of Social and Behavioral Sciences at Utrecht University.

Design. To reiterate, in a two-armed bandit task with two bags, A and B, and two outcomes, blue and yellow balls, partici-

² For the main effect of preferring one button over the other resulting in a pseudocontingency. Assuming this preference drives sampling, this measure is most closely related to our research question.

pants encountered initial evidence before then sampling freely. One of the two colors would result in the participant winning 10 points while the other color would result in the participant losing 10 points. Participants were randomly assigned to either of two conditions, a reward-rich or a reward-impooverished condition. We counterbalanced which bag would be more frequently shown in the initial evidence phase, which color was the rewarding one, and which color participants were asked about when giving estimates of relative contingency and conditional probability estimates. The experiment is therefore a simple two-factor between-participants design.

Procedure. The experiment was divided into three phases: An induction phase, in which participants were guided to choose particular options that made up the distribution of initial evidence. A free sampling phase, in which participants could choose freely between both options. And the final phase, in which participants gave estimates regarding the task. The experiment lasted a total of 100 trials and participants were told to maximize their rewards in a gambling task. Participants were informed that they would sample with replacement, so that the overall number of blue and yellow balls would remain constant. This was done so as to deter participants from trying to detect patterns in the environment but instead focus on the outcome probabilities without expecting them to change throughout the experiment. The points earned were tallied and determined their incentivized payoff.

In the induction phase, participants were told that the computer would randomly determine which bag was to be chosen so as to get participants familiar with the task. To this end, they would then only see the one available option. For example, they would see Bag A, and, after clicking on this option and a short delay, receive feedback in the form of text (“You chose bag A and drew a yellow ball.”) as well as images depicting in this case Bag A and a yellow ball with “+10” (or “−10”) written on the ball. After a delay of 1 s the feedback would disappear, and the next choice was presented. During the entire experiment, the current trial number as well as the total trial number (“Trial: x/100”) were presented on the screen thereby making the horizon of the sampling phase salient and allowing participants to know the remaining number of trials to either explore or exploit.

The induction phase consisted of four trials that introduced the initial evidence used to induce a bias toward one of the two bags. In the reward-rich condition, participants won on all three trials in which they had to select the one bag (e.g., Bag A resulted in yellow balls with a +10 being drawn) while they lost on the one trial in which they had to select the other bag (e.g., Bag B resulted in a blue ball with a −10 being drawn). In the reward-impooverished condition, participants lost on all three trials in which they had to select the one bag (e.g., Bag B, −10 blue ball) while they won on the one trial in which they had to select the other bag (e.g., Bag A, +10 yellow ball). To summarize, the resulting sampling pattern of the two options was A-A-B-A for the first four trials with the outcome pattern being either win-win-loss-win or loss-loss-win-loss depending on whether participants were in the reward-rich or reward-impooverished condition, respectively (see Table 1).

In the free sampling phase, participants were free to choose either of the bags on any given trial for the remaining 96 trials of the experiment. While we manipulated the distribution of outcomes during the induction phase trials, during the sampling phase the outcomes were randomly drawn from the remaining list of

outcomes (out of 100) for each location. Hence, across all 100 trials the two options were exactly equal. Our behavioral measure was the number of choices participants made from the two bags during this free sampling phase.

In the final phase, participants gave estimates regarding their relative contingency estimates between the two bags and the outcomes. Specifically, we asked them from which of the two bags they were more likely to grab a blue (yellow) ball. Participants answered by moving a slider which was anchored with the two bags displayed as images at the ends. We refer to this measure as relative contingency estimate. Participants also gave conditional probability estimates for both bags. We asked them how likely it was to grab a blue (yellow) ball if they chose Bag A (and Bag B). They again answered by moving a slider, this time anchored at 0% and 100%. Additionally, we asked participants to indicate how confident they were in making a reasonable estimate regarding each bag. This slider was anchored at *not confident at all* and *very confident*. Except for the relative contingency estimate, all scales had an indication of the slider marker’s current position in percent that was updated as the slider was moved. We counterbalanced the color asked for in the dependent variables (DVs) so as to ask for the frequent or infrequent, winning or losing color. In other words, if blue (yellow) balls were frequently shown, we counterbalanced whether participants were asked to give estimates regarding the blue or the yellow balls.

As there was a large discrepancy in how often participants could win between conditions, the payoffs were determined per condition. The points earned during the task were transformed so that the minimum (maximum) number of points would represent earning the minimum (maximum) payoff and the average number of points earned would equal the average payoff. The payoff was therefore relative to the peers of one’s group.

Data preparation. All slider values were later transformed and recoded such that 0 would always represent neutrality between both options and positive values represent the option that was encountered more frequently in the induction phase. That is, if Bag A was shown three times (three yellow +10 balls being drawn in the reward-rich or three blue −10 balls being drawn in the reward-impooverished condition), all measures for this participant were recoded such that a positive value on the relative contingency estimate measure would indicate a bias toward Bag A and a positive value on the conditional probability estimates would indicate a positive contingency between Bag A and the frequent, yellow balls.

We also report post hoc results of a binomial test for the relative contingency estimate that hopefully reduces some ambiguity as to the exact interpretation of the slider.³ We deem a binomial test to be the best choice as we are confident that values above or below zero indicate a preference for the respective option. We formed binary groups based on participants’ relative contingency estimate, excluding any with a score of exactly zero. We then expected a proportion of participants larger than .5 to have preferred the more frequent bag in the reward-rich condition, but

³ After data collection, we realized that there might be some ambiguity as to how the relative preference estimate measure should be interpreted by participants as we had failed to make entirely clear whether a more extreme measure should indicate stronger preference or higher confidence.

that the proportion would not be different from .5 in the reward-impooverished condition.

Data preparation and analyses was undertaken using R (R Core Team, 2018) and especially the packages “dplyr” (Wickham, François, Henry, & Müller, 2019), “BayesFactor” (Morey & Rouder, 2018), “lme4” (Bates, Maechler, Bolker, & Walker, 2015), and “lmerTest” (Kuznetsova, Brockhoff, & Christensen, 2017). Across all three measures of preference—the sampling choices made, the relative contingency estimates, and the conditional probability estimates—we expect a bias in the reward-rich condition and therefore perform one-tailed tests reporting Bayes factors (BF_{+0}). We expect no bias in the reward-impooverished condition and therefore perform equality tests (BF_{01}). And finally, we test the difference between conditions by directly comparing the deviation from chance level between both conditions and expect this deviation to be larger in the reward-rich than in the reward-impooverished condition, which again warrants one-tailed testing (BF_{+0}). While we predict attenuation in the reward-impooverished condition, the mean bias of participants may fluctuate slightly around chance level. For example, we expect relative contingencies to have attenuated toward chance level in the reward-impooverished condition. But participants in this condition might still have a slight preference for the option they encountered less often (and associate less with losing), while participants in the reward-rich condition should prefer the option they encountered more often (and associate with more winning). While for most analyses we recoded data by frequency of options, we use the absolute deviation from chance level to compare the two conditions (see also Nieuwenhuis, Forstmann, & Wagenmakers, 2011). We refer to this test as the strength of the effect to clearly indicate this comparison. The Bayes factor quantifies the likelihood of the data to be observed under one hypothesis compared to a competing hypothesis. The subscript indicates the direction of the comparison such that, for example, BF_{01} indicates the relative support for H_0 over the competing hypothesis H_1 (Hojtink, Mulder, van Lissa, & Gu, 2019). All Bayesian tests use the default prior of the Bayes-Factor package, namely a Cauchy distribution of width $r = \frac{\sqrt{2}}{2}$. In the Supplemental Material A, we include further Bayesian analyses in the form of 95% highest density intervals, the median of this interval as a Bayesian effect size estimate, as well as robustness checks.

To analyze the choice behavior, we coded every choice of the frequent bag as +1 and every choice of the alternative as 0, effectively creating a choice index of participants’ overall preference. All graphs include confidence intervals around the means, and we report confidence intervals for the effect sizes.

Results

Sampling. Over the 96 trials of the free sampling phase, participants in the reward-rich condition sampled the frequent bag, that is the bag that was shown three times with rewarding outcomes during the induction phase, on average 57% ($SD = 19.26$) of the time. In the reward-impooverished condition on the other hand, participants sampled the frequent (three losses) bag on average 46% ($SD = 14.95$) of the time. This is above chance in the reward-rich condition, $BF_{+0} = 5.34$, $t(51) = 2.53$, $p = .007$, $d = .35$, 95% $CI_d [0.07, 0.63]$, and some tentative indication that participants were not different from chance in the reward-

impooverished condition, $BF_{01} = 1.49$, $t(47) = -1.78$, $p = .082$, $d = -0.26$, 95% $CI_d [-0.54, 0.03]$. However, the difference in the strength of the biases (the comparison of the respective difference of the two conditions from chance level, see Data Preparation section) was not significant, $BF_{+0} = 0.46$, $t(95.26) = 0.85$, $p = .198$, $d = 0.17$, 95% $CI_d [-0.23, 0.57]$. See also Figures 2 and 3.

We then analyzed participants’ choices over time (trials) by means of a general mixed model in which participants were entered as random effect and trial number, condition, and the interaction were treated as fixed effects. We scaled and shifted trial number so that the first free choice trial would be at timepoint zero in the model. Due to the binary outcomes, we fitted the following logistic model⁴ to the data in which the reward-impooverished condition is our control and trial number is scaled:

$$\hat{y} = c + \frac{trial}{100} \cdot \beta_{trial} + condition \cdot \beta_{condition} + \frac{trial}{100} \cdot condition \cdot \beta_{trial*condition} \quad (3)$$

First, the intercept was not significant which can be seen as testament to how quickly participants in the reward-impooverished condition attenuated to chance-level, $c = -0.24$, $z = -1.52$, $p = .129$. Importantly though, there was a significant main effect of condition indicating a difference between conditions and the successful induction of a bias in the reward-rich condition, $\beta_{condition} = 0.88$, $z = 4.01$, $p < .001$. Next, the positive estimate for trial number indicates attenuation toward chance-level in the reward-impooverished condition, $\beta_{trial} = 0.18$, $z = 1.61$, $p = .107$. The negative interaction term indicates that the difference between conditions decreases over time, $\beta_{trial*condition} = -0.72$, $z = 4.54$, $p < .001$. Finally, we tested for biases on the last trial by transforming trial number such that the last trial would be the null point in the model. As expected, the model indicated full attenuation in the reward-impooverished condition as indicated by the nonsignificant intercept, $c = -0.06$, $z = -0.37$, $p = .709$. However, participants in the reward-rich condition seem to have also attenuated according to this model as indicated by the nonsignificant main effect of condition, $\beta_{condition} = 0.16$, $z = 0.75$, $p = .457$.

We hypothesized that the attenuation in the reward-impooverished condition might be due to higher oscillation between the two bags. This was, however, not the case as the average percentage of participants in the reward-impooverished condition that switched was only slightly higher ($M = 34.31$, $SD = 8.61$) than participants in the reward-rich condition ($M = 33.77$, $SD = 6.09$), $BF_{+0} = 0.24$, $t(171.09) = 0.5$, $p = .31$, $d = 0.07$, 95% $CI_d [-0.21, 0.36]$. Noteworthy, however, is that following the induction phase 82.7% of participants in the reward-rich condition opted to select the option frequently shown during the induction phase, but only 10.4% of participants in the reward-impooverished condition. The remaining trials differ less drastically as can be seen in Figure 4. While we would predict participants to be inclined to explore both options as sampling continues (e.g., to alleviate boredom; Mehlhorn et al., 2015), there is little difference detectable even in early trials.

⁴ We also fitted quadratic models. However, the quadratic model only outperformed the linear model in Experiment 1 and then not by much. For reasons of comparison, we opted to always use the linear model.

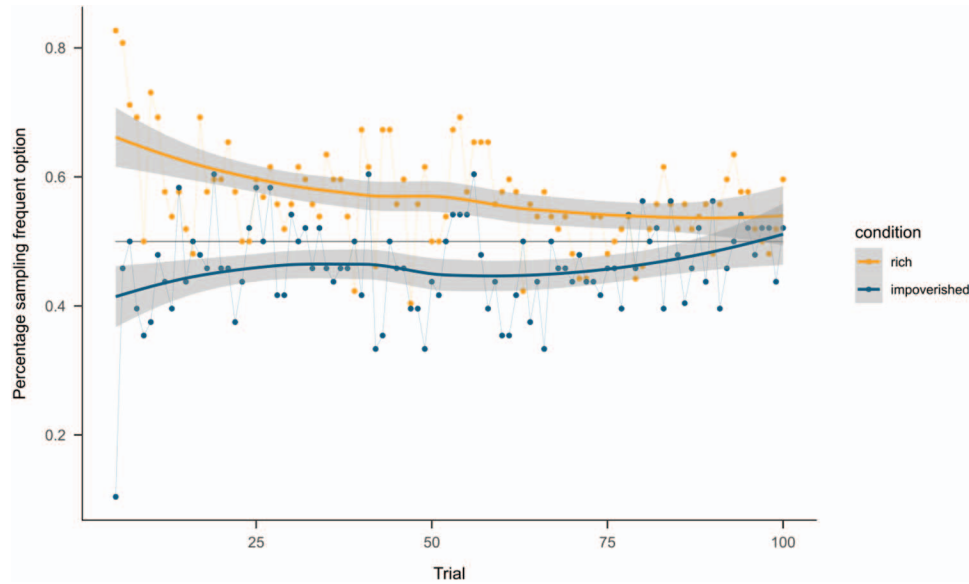


Figure 2. Percentage of participants sampling the frequent option per trial (Experiment 1). See the online article for the color version of this figure.

Relative contingency estimates. The second preference measure was the relative contingency estimate people made after finishing the sampling. In the reward-rich condition, the effect failed to reach significance ($M = 7.15$, $SD = 31.81$), $BF_{+0} = 0.96$, $t(51) = 1.62$, $p = .056$, $d = 0.22$, 95% $CI_d [-0.05, 0.50]$. In the reward-impoverished condition, the data, as expected, supported the absence of any bias ($M = 0.29$, $SD = 28.79$), $BF_{01} = 6.36$, $t(47) = 0.07$, $p = .944$, $d = 0.01$, 95% $CI_d [-0.27, 0.29]$. Given the absence of an effect in the reward-rich condition it will come as no surprise that there was no support for a difference between conditions when testing the strength of the effect, $BF_{+0} = 0.63$, $t(97.96) = 1.13$, $p = .13$, $d = 0.23$, 95% $CI_d [-0.17, 0.62]$.

Because the labeling may have been unclear and participants could have understood our relative contingency scale as extreme values referring to either stronger contingency or higher confidence, we also analyzed the responses in binary format and ran binominal tests. Six participants were excluded from these binominal tests because their score was exactly zero (see also Data Preparation section). The binominal tests indicated that in the reward-rich condition the proportion (32 classified as preferring the frequent bag out of the 49 participants in this condition) was over three times more likely to be larger than chance, $BF_{+0} = 3.41$; $p = .022$. In the reward-impoverished condition on the other hand the proportion (23/45) was over five

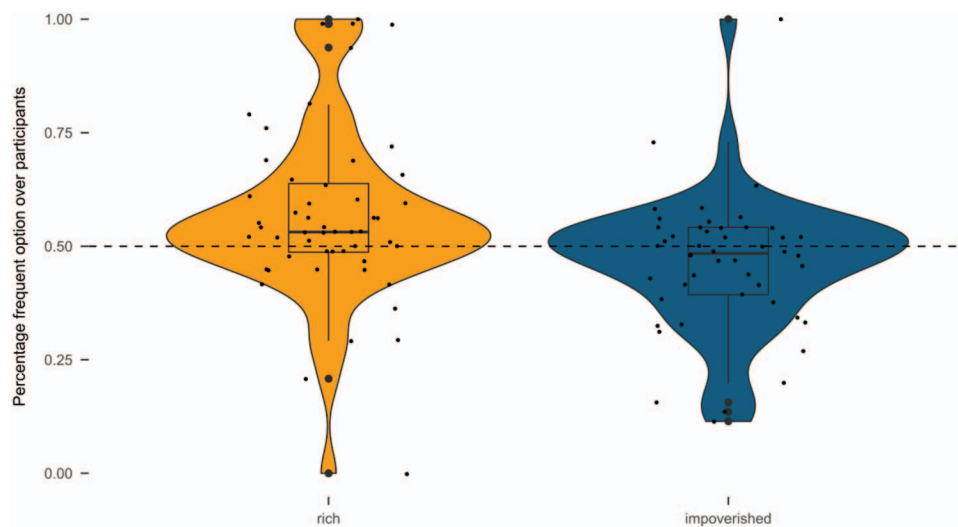


Figure 3. Proportion of choosing the frequent option per participant (Experiment 1). See the online article for the color version of this figure.

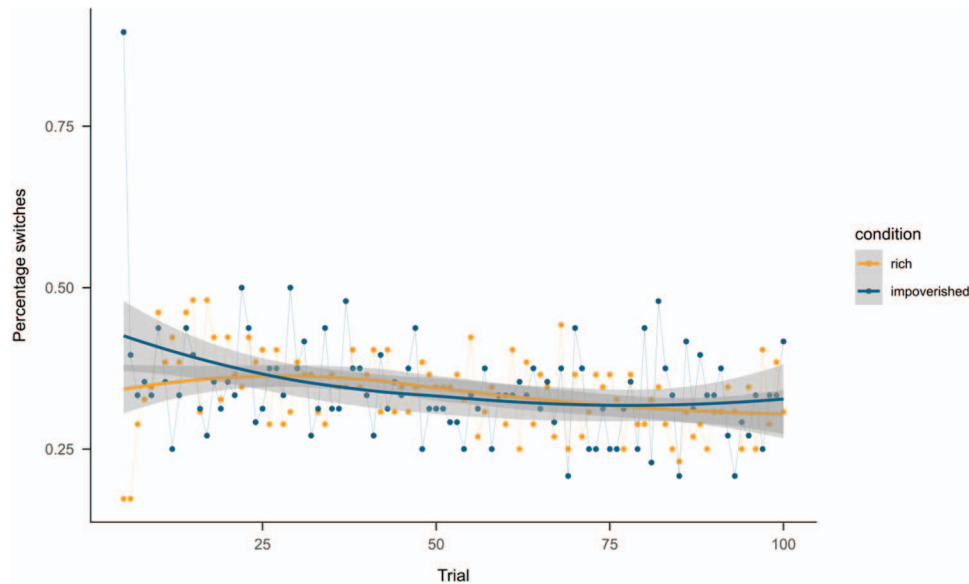


Figure 4. Percentage of participants switching from previous choice including trend lines per condition (Experiment 1). See the online article for the color version of this figure.

times more likely to reflect chance-level behavior, $BF_{01} = 5.38$; $p = .617$. To test the strength of the effect, we then tested the proportion in the reward-rich against the proportion in the reward-impoverished condition finding support for a difference between the two conditions, $BF_{+0} = 4.86$; $p = .015$.

Conditional probability estimates. The third preference measure were the conditional probability estimates for which participants gave estimations of how likely they were to grab a blue (or yellow, depending on the counterbalancing) ball given that they had chosen Bag A and Bag B, respectively. In the reward-rich condition, participants estimated that they drew the frequent (winning) color 64.65% ($SD = 19.77$) of the time when choosing the frequent bag, but only 52.31% ($SD = 20.30$) of the time when choosing the infrequent bag. In the reward-impoverished condition, participants estimated that they drew the frequent (losing) color 66.35% ($SD = 22.42$) of the time when choosing the frequent bag and 66.17% ($SD = 22.00$) of the time when choosing the infrequent bag. From these estimations we then calculated difference scores between the two conditional probabilities in the form of ΔP -scores (Allan, 1980). The mean ΔP -score in the reward-rich condition was $\Delta P = .12$ ($SD = 0.32$), which differs from 0, $BF_{+0} = 10.05$, $t(51) = 2.81$, $p = .003$, $d = 0.39$, 95% CI_d [0.11, 0.67], indicating relatively higher estimates of the rewards from the frequent bag. In the reward-impoverished condition, the mean ΔP -score as expected did not differ from 0, $\Delta P = .00$ ($SD = 0.25$), $BF_{01} = 6.37$, $t(47) = 0.05$, $p = .959$, $d = 0.01$, 95% CI_d [-0.28, 0.29]. Testing the strength of the effect, the data holds evidence that the reward-rich condition differs from the reward-impoverished condition, $BF_{+0} = 2.95$, $t(95.83) = 2.14$, $p = .018$, $d = 0.42$, 95% CI_d [0.02, 0.82].

Confidence. Finally, participants' confidence in their judgments did not differ between the frequent ($\Delta P = -.01$, $SD = 0.23$) and the infrequent option ($\Delta P = .03$, $SD = 0.20$), $BF_{01} = 6.35$, $t(197.45) = 0.23$, $p = .822$, $d = 0.03$, 95% CI_d [-0.25, 0.31].

Discussion

The aim of this study was to validate the experimental design and compare sampling choices and preference biases in a reward-rich and a reward-impoverished environment as a between-subjects factor. In summary, the results provide some support for our hypothesis. Specifically, biased preferences in behavioral choices as well as in subsequent estimates were obtained in the reward-rich but not the reward-impoverished condition. The results highlight the influence of the environment on the process of updating beliefs and attest to the strong influence of initial evidence. It seems that indeed the environment decision makers find themselves in drastically changes how they balance the trade-off between exploration and exploitation. That being said, the results are not as clear-cut as the simulations seem to suggest. One explanation for this might be that the extremity of the initial evidence distribution not only allowed for a strong initial belief, it also allowed, at least implicitly, for very easy falsification as the outcomes had no variability (outcomes were either wins or losses without gradation). It is quite possible that participants in the reward-rich condition, upon encountering new evidence in the free sampling phase, hesitated to follow an exploitation scheme. Instead, they may have opted for an extended exploration period after the outcomes of the free sampling phase proved not to live up to the expectations raised by the initial evidence. Interestingly, however, participants did still indicate biases when explicitly asked to estimate the conditional probabilities.

Another interesting finding is the absence of any difference in confidence participants have in their estimates regarding the frequent and the infrequent bag. This finding is perfectly in line with a base rate alignment account such as pseudocontingencies, as the underlying mechanism would predict people to associate the frequent action with the frequent outcome just as strongly as the infrequent action with the infrequent outcome. While BIAS and

MDM do not specifically mention confidence (a feature that certainly could be implemented in either model), a Bayesian account would clearly predict higher confidence in the more frequent alternative (with the probability density function being narrower and with a higher maximum around the best estimate).

In sum, participants only partially exhibit the pattern we predicted, which may be due to the unrealistically positive or negative manipulation of initial evidence. Experiment 2 therefore aims to extend the findings of Experiment 1 by inducing the bias via a distribution which is representative of the overall distribution and without a contingency proper between locations and outcomes.

Experiment 2a: No Contingency Initial Evidence

In the second study, the initial evidence consisted of a distribution meant to elicit pseudocontingencies without containing a contingency proper. We decided to use this distribution, first, because it represented the evidence (i.e., only base rates) participants would encounter during the sampling phase much better. The proportion of wins to losses was exactly the proportion participants would encounter throughout the entire experiment. And second, this distribution allowed us to induce biases despite the actual contingency being zero. Again, this is in line with the actual setup of the experiment in which neither option is better than the alternative. But it also makes more apparent the notion that unrealistic biases are maintained.

Let us therefore briefly revisit pseudocontingencies. Instead of relying on the pairwise occurrences, a pseudocontingency inference is formed by aligning the skewed base rates of cues and outcomes and linking the frequent cue with the frequent outcome and the infrequent cue with the infrequent outcome (Fiedler & Freytag, 2004; Fiedler et al., 2009). Interestingly, pseudocontingencies have been shown to also influence choice behavior (Meiser et al., 2018) and override genuine contingencies (Fiedler, 2010). We therefore expected participants to form biases after encountering the initial distribution despite there being no genuine contingency.

Additionally, in order to investigate the influence of the initial evidence and of the sampling phase separately, Experiment 2a had participants indicate their preference and estimates twice. First, immediately following the induction phase and, second, as in the previous study, after the free sampling phase.

In order to ensure maximal attention to all outcomes, the payoffs for the balls were only introduced after the initial sampling phase. That is, participants were presented with the guided samples of the induction phase, then gave their initial estimates, and only then learned which color would earn them points and which one would cost them points in the following free sampling phase. During the free sampling phase, they still received feedback in that the counter in the corner of the screen depicted the current point tally. Would participants form erroneous initial beliefs and then maintain them in the reward-rich condition even though throughout the entire experiment the actual contingency remained zero? If the initial evidence is successful in inducing pseudocontingencies, participants should initially favor the frequent option in both conditions. Then, as they sample freely, the initial bias should be perpetuated in the reward-rich but attenuated in the reward-impovertished condition.

Method

We decided to double the sample size compared with Experiment 1 so as to rule out any potential problems concerning statistical power. Two-hundred participants ($N_{\text{female}} = 122$) with an average age of 33 years ($SD = 9.65$) participated for a financial reward of £0.85 plus additional earnings (max £2.15, mean £1.50) based on performance. Participants for this study were again recruited via the online platform Prolific Academic and the study was run in English on Soscisurvey (Leiner, 2014).

This experiment builds on Experiment 1 with three important changes: First, we use a different distribution for the induction phase which has a contingency proper of zero (see Table 1). The order of the 16 trials that make up the initial evidence was random. Second, the dependent variables were asked twice, once after the initial evidence phase and once, as before, after the sampling phase. And third, rewards were only introduced after the initial evidence phase. That is, participants were introduced to the two bags and the two colors that could be drawn from the bags, underwent the induction phase, and answered the DVs a first time. Only then were \pm points introduced and linked to ball colors. A few minor changes were necessary to accomplish this: Throughout the entire experiment the balls no longer had text on them stating whether them being drawn resulted in a win or a loss (see Table 1). Furthermore, the feedback no longer contained information about whether the participant had won or lost 10 points on that trial. These changes were made so as to keep the colors neutral during the initial evidence phase while also keeping the feedback consistent across all trials. As in the last experiment, all variables were transformed and recoded such that positive values represent a preference for the frequent option.

Results

Relative contingency estimates premeasure. After the induction phase and before we introduced the reward-scheme, we expected participants in both conditions to associate the frequent bag with the frequent outcome. In other words, we expected both conditions to have developed a pseudocontingency inference, as the balls were still neutral at this point.

Indeed, participants developed the same biases in both conditions. There was no difference between conditions, $BF_{01} = 3.47$, $t(197.88) = 1.16$, $p = .247$, $d = 0.16$, 95% $CI_d [-0.12, 0.44]$. But participants across conditions exhibited a bias toward the frequent bag, ($M = 12.84$, $SD = 31.95$), $BF_{+0} = 411,159.82$, $t(199) = 5.68$, $p < .001$, $d = 0.40$, 95% $CI_d [0.26, 0.55]$.

Two participants were excluded from the binary analyses due to their relative contingency equaling exactly zero. In the reward-rich condition, 72 participants classified as preferring the more frequently shown bag out of the 102 participants in this condition. In the reward-impovertished condition, 63 classified as preferring the more frequently shown bag out of the 96 participants. There was again no difference between the two conditions, $BF_{01} = 4.24$, $p = .171$, but a strong overall bias such that more participants were classified as preferring the frequent bag, $BF_{+0} = 54,343.30$, $p < .001$.

Conditional probability estimates premeasure. Participants estimated that they drew the frequent color 64.23% ($SD = 21.48$) of the time when choosing the frequent bag but only 51.40% ($SD = 23.00$) of the time when choosing the infrequent bag. The

mean ΔP -score in the reward-rich ($\Delta P = .12$, $SD = 0.35$) and reward-impooverished ($\Delta P = .14$, $SD = 0.38$) condition did not differ from one another, $BF_{01} = 5.93$, $t(195.28) = -0.44$, $p = .34$, $d = -0.06$, 95% $CI_d [-0.34, 0.22]$, and suggest a large overall effect, $BF_{+0} = 16,802.19$, $t(199) = 4.99$, $p < .001$, $d = 0.35$, 95% $CI_d [0.21, 0.50]$.

Confidence premeasure. Participants estimated their confidence similarly in the reward-rich ($\Delta P = .05$, $SD = 0.23$) and the reward-impooverished condition ($\Delta P = -.03$, $SD = 0.23$), $BF_{01} = 8.31$, $t(397.88) = 0.42$, $p = .677$, $d = 0.04$, 95% $CI_d [-0.15, 0.24]$.

Sampling. Replicating the findings from Experiment 1, participants again showed biases in sampling in the reward-rich condition but attenuated any biases in the reward-impooverished condition as can be seen in Figures 5 and 6. Over the 84 trials of the free sampling phase, participants in the reward-rich condition sampled the frequent bag on average 60% ($SD = 27.35$) of the time. In the reward-impooverished condition on the other hand, participants sampled the frequent bag on average 48% ($SD = 18.65$) of the time. In other words, we find a bias in the reward-rich condition, $BF_{+0} = 124.08$, $t(102) = 3.72$, $p < .001$, $d = 0.37$, 95% $CI_d [0.17, 0.57]$. We find that participants do not differ from chance-level in the reward-impooverished condition, $BF_{01} = 5.99$, $t(96) = -0.90$, $p = .368$, $d = -0.09$, 95% $CI_d [-0.29, 0.11]$. Further, the strength of the bias differed between the two conditions, $BF_{+0} = 5.53$, $t(180.78) = 2.53$, $p = .006$, $d = 0.35$, 95% $CI_d [0.07, 0.63]$.

We then analyzed participants' choices over time (trials) by means of the same mixed linear model as in Experiment 1 (with participants as random effect and trial number, condition, and their interaction as fixed effects; see also Formula 3). As in Experiment 1, the nonsignificant intercept indicates how quickly participants in the reward-impooverished condition attenuated to chance level ($c = -0.13$, $z = -0.79$, $p = .430$) and the main effect for

condition conversely indicates a successful bias induction in the reward-rich condition, $\beta_{condition} = 0.97$, $z = 4.09$, $p < .001$. The positive estimate for trial number indicates attenuation in the reward-impooverished condition, $\beta_{trial} = 0.10$, $z = 1.04$, $p = .299$. The interaction term again indicates the difference between conditions to decrease over time $\beta_{trial \times condition} = -0.45$, $z = -3.10$, $p = .002$. Once again shifting the model to analyze the last trial, we found attenuation in the reward-impooverished condition ($c = -0.03$, $z = -0.18$, $p = .861$), but a main effect for condition indicating a significant difference between conditions and a persisting bias in the reward-rich condition, $\beta_{condition} = 0.52$, $z = 2.12$, $p = .034$.

Interestingly, we now also found the predicted effects on oscillation between the two options. Participants in the reward-impooverished condition switched more often ($M = 30.0$, $SD = 6.02$) than participants in the reward-rich condition ($M = 25.0$, $SD = 4.55$), $BF_{+0} > 1,000,000.00$, $t(154.55) = 6.11$, $p < .001$, $d = 0.94$, 95% $CI_d [0.62, 1.26]$. Compared with the pattern we found in Experiment 1, we found a less pronounced difference on the first trial as 58.8% of participants in the reward-impooverished and 29.1% of participants in the reward-rich condition switched, as can be seen in Figure 7.

Relative contingency estimates postmeasure. After the introduction of the reward-scheme and participants engaging in the free sampling phase, we expected participants to have maintained their initial bias in the reward-rich, but to have attenuated the bias in the reward-impooverished condition. Indeed as can also be seen in Figure 8, participants exhibited a bias in the reward-rich condition ($M = 9.00$, $SD = 32.05$), $BF_{+0} = 9.82$, $t(102) = 2.85$, $p = .003$, $d = 0.28$, 95% $CI_d [0.08, 0.48]$, but not in the reward-impooverished condition, demonstrating in the expected null effect ($M = -3.64$, $SD = 26.87$), $BF_{01} = 3.78$, $t(96) = -1.33$, $p = .185$, $d = -0.14$, 95% $CI_d [-0.34, 0.06]$. While, in contrast to Experiment 1, we now do find a significant effect in the reward-rich

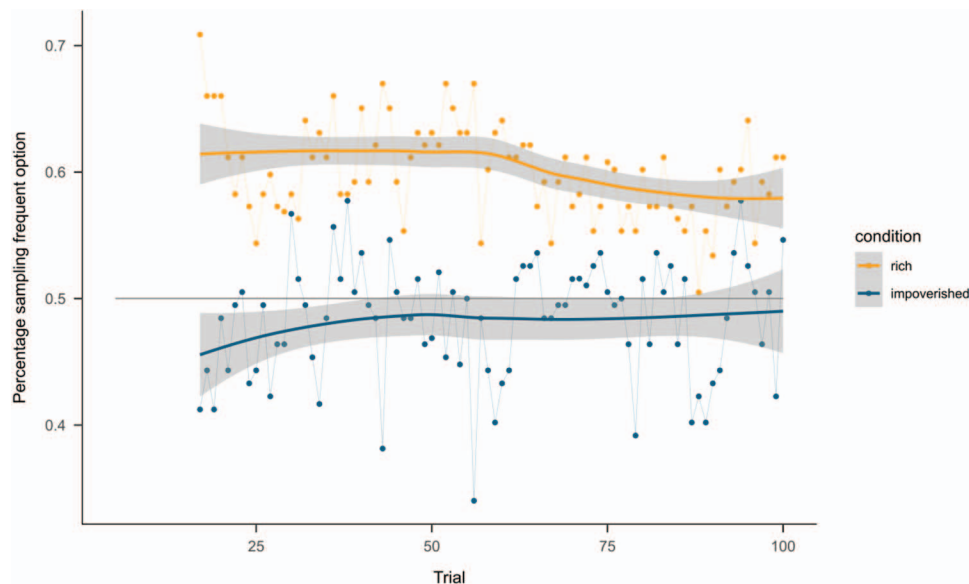


Figure 5. Percentage of participants sampling the frequent option per trial (Experiment 2a). See the online article for the color version of this figure.

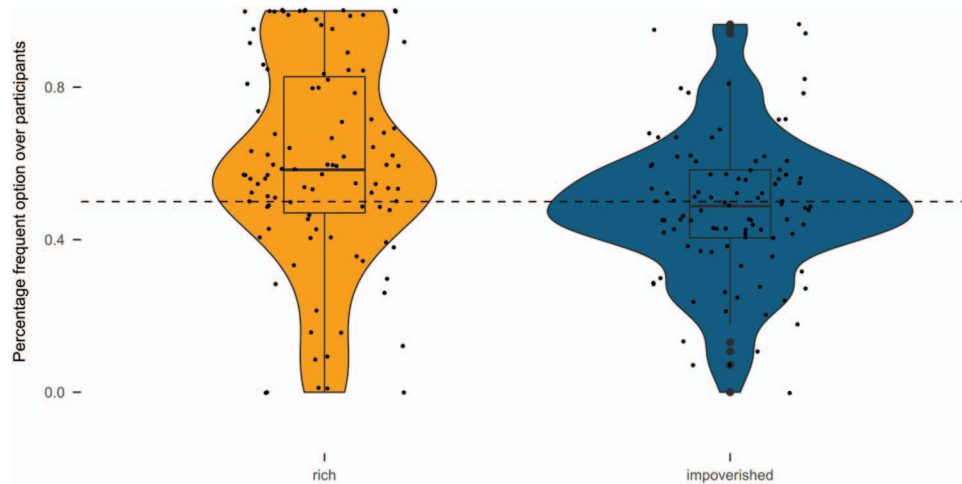


Figure 6. Proportion of choosing the frequent option per participant (Experiment 2a). See the online article for the color version of this figure.

condition, we still found no difference between both conditions when testing the strength of the effect, $BF_{+0} = 0.59$, $t(195.41) = 1.28$, $p = .100$, $d = 0.18$, 95% $CI_d [-0.10, 0.46]$.

In our binominal analyses, we replicate the pattern we found in Experiment 1. Three participants were excluded from the binary analyses due to their relative contingency equaling exactly zero. In the reward-rich condition 66 participants preferred the frequent bag out of the 101 participants in this condition, $BF_{+0} = 29.64$, $p = .001$. In the reward-impoverished condition 46 out of the 96 classified as preferring the frequent bag which does not differ from chance-level behavior, $BF_{01} = 7.25$, $p = .695$. To test the strength of the effect, we then tested the proportion in the reward-rich against the proportion in the reward-impoverished condition find-

ing support for a difference between the two conditions, $BF_{+0} = 8.73$, $p = .005$.

Conditional probability estimates postmeasure. After sampling, participants' estimates for grabbing a ball of the frequent (winning) color in the reward-rich condition (frequent bag: 62.38%, $SD = 21.08$, infrequent bag: 54.93%, $SD = 21.74$) resulted in a mean ΔP -score of $\Delta P = .07$ ($SD = 0.34$). This is about two times more likely to represent the predicted effect of a pseudocontingency inference in favor of the frequent option, $BF_{+0} = 2.14$, $t(102) = 2.2$, $p = .015$, $d = 0.22$, 95% $CI_d [0.02, 0.41]$. Participants' estimates in the reward-impoverished condition (frequent, losing ball: frequent bag: 69.66%, $SD = 20.90$, infrequent bag: 71.02%, $SD = 22.04$) resulted in a mean ΔP -score of $\Delta P = -.01$ ($SD = 0.25$). This is more likely

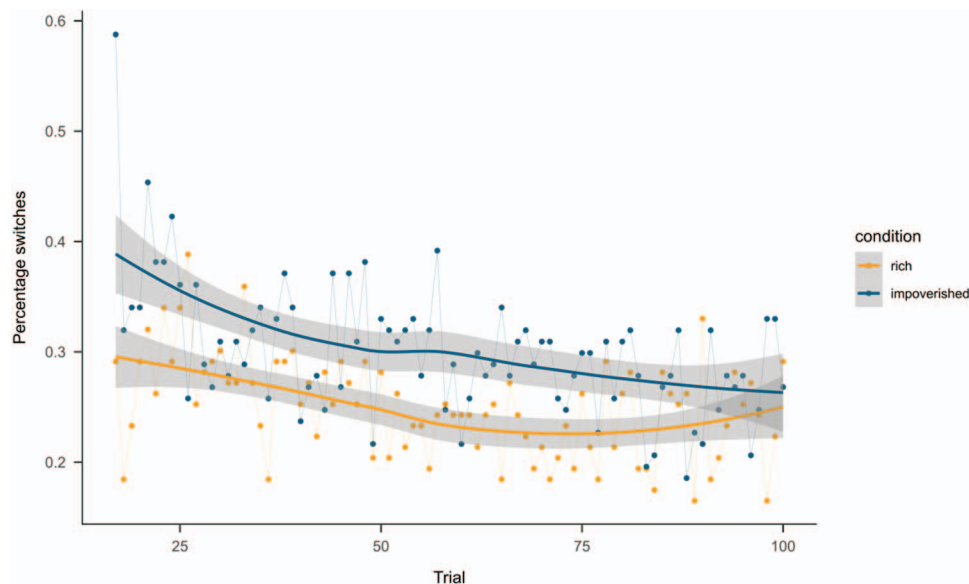


Figure 7. Percentage of participants switching from previous choice including trend lines per condition (Experiment 2a). See the online article for the color version of this figure.

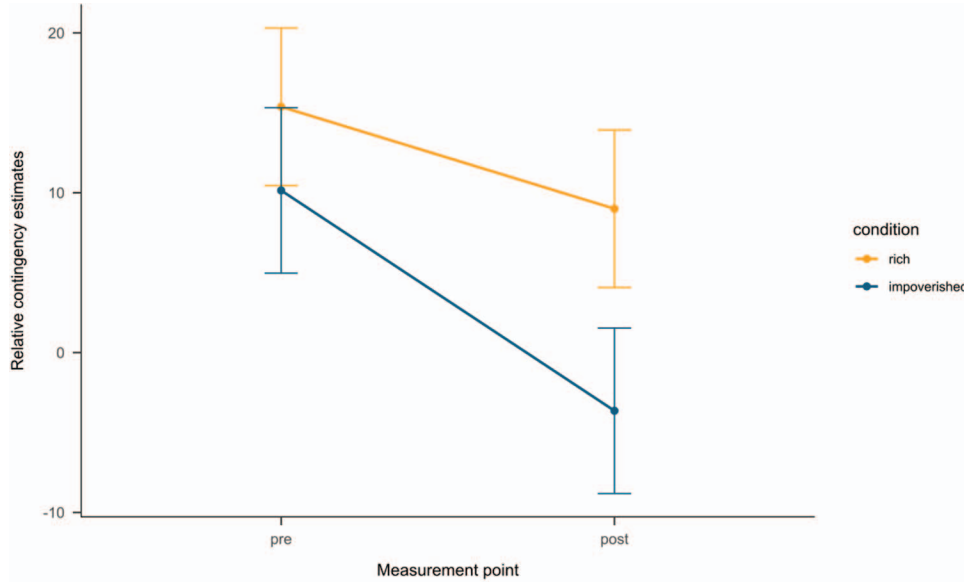


Figure 8. Relative contingency estimates for the pre and post measurement for the frequent over the infrequent bag (Experiment 2a). See the online article for the color version of this figure.

to reflect a null effect, $BF_{01} = 7.74$, $t(96) = -0.54$, $p = .593$, $d = -0.05$, 95% $CI_d [-0.25, 0.14]$. There was once again support for the strength of the effect, $BF_{+0} = 2.17$, $t(186.27) = 2.08$, $p = .019$, $d = 0.29$, 95% $CI_d [0.01, 0.57]$. In summary, we replicate the pattern found in Experiment 1 (see also Figure 9).

Confidence postmeasure. Finally, participants again did not estimate their confidence in their judgments to be different in the reward-rich ($\Delta P = -.02$, $SD = 0.25$) or the reward-infrequent ($\Delta P = -.02$, $SD = 0.19$) condition, $BF_{01} = 7.37$,

$t(397.82) = -0.65$, $p = .516$, $d = -0.07$, 95% $CI_d [-0.26, 0.13]$.

Discussion

The aim of Experiment 2a was to replicate Experiment 1 using a distribution that contained a contingency of zero but was still likely to bias participants. The results from the premeasurement across all measures speak to the successful induction of a bias by

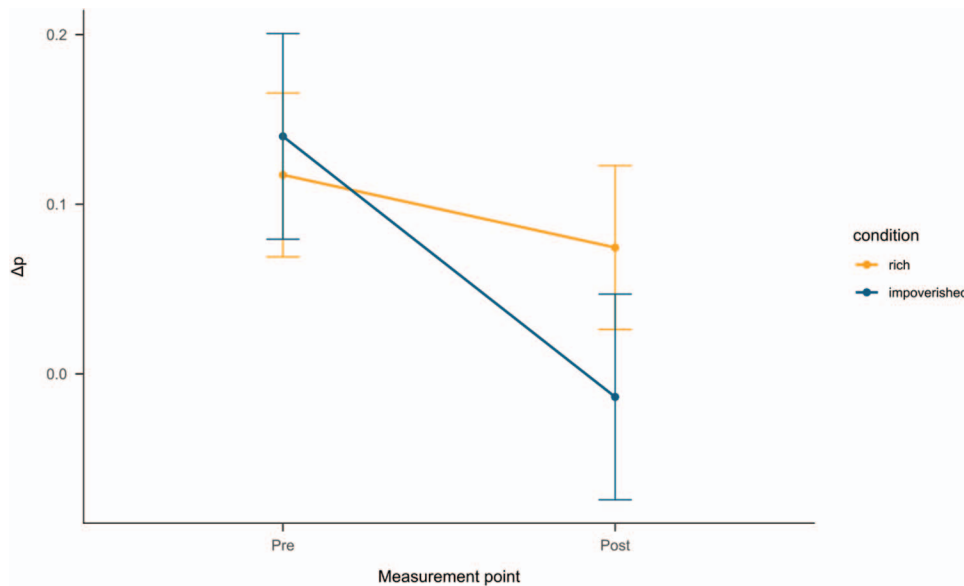


Figure 9. ΔP-scores from conditional estimates (Experiment 2a). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the initial evidence and serve as a manipulation check for the pseudocontingency manipulation. Thereafter, we introduced rewards and losses hypothesizing that participants would maintain their biases in the reward-rich condition but attenuate them in the reward-impooverished condition. The results across the three preference measures at the second measurement time support this notion. In summary then, across two experiments we found fairly consistent support for the maintenance of a bias in the reward-rich condition both during sampling and afterward in relative contingency and conditional probability estimates. We found consistent support for the attenuation of any biases in the reward-impooverished condition. And we found mixed results regarding the strength of the effect, that is, the contrast between conditions.

Why do the results show a clearer pattern than the first experiment? As already outlined above, we believe the unrealistically positive or negative outcomes in the initial evidence to be the cause. While Experiment 1 had a distribution that was easily confirmed or falsified thereafter (cf. Pilditch & Custers, 2018), Experiment 2a utilized a distribution that was representative of the overall distribution. Specifically, the proportion of wins and losses for both options was identical to the proportion the options had over the entirety of the experiment. Next, in Experiment 2b, we replicated these findings in the lab.

Experiment 2b: Replication

Method

This study was run in the lab at Utrecht University as a filler task of another, unrelated study for which they received monetary compensation. One hundred and ninety-eight ($N_{\text{female}} = 143$) participated in this study. Following the induction phase, participants were once again informed about the points they could earn or lose with balls in the two respective colors. The top 25% of participants for each condition (we used a cutoff based on previous experiments that was unknown to participants) could win an additional 1€ on top of their regular payment for participation. The experiment is identical to Experiment 2a other than that for both the pre- and the postmeasure, participants were additionally asked to indicate base rate estimates after estimating their relative contingency and before estimating conditional probability estimates. The base rate estimates allow for an estimation of the perceived subjective skewness of options chosen and the outcomes thereof. We expect higher subjective skewness to go hand in hand with stronger maintained biases.

Results

Relative contingency estimate premeasure. Replicating our findings from Experiment 2a, we found no difference between conditions following the induction phase (and before they learned which color would be rewarding), $BF_{01} = 5.28$, $t(195.73) = 0.66$, $p = .511$, $d = 0.09$, 95% CI_d [-0.19, 0.37]. Participants across conditions exhibited a bias toward the frequent bag ($M = 10.83$, $SD = 25.48$), $BF_{+0} > 1,000,000.00$, $t(197) = 5.98$, $p < .001$, $d = 0.43$, 95% CI_d [0.28, 0.57].

Four participants were excluded from the binary analyses due to their relative contingency equaling exactly zero. In the reward-rich condition, 68 participants classified as preferring the more fre-

quently shown bag out of the 97 participants in this condition. In the reward-impooverished condition, 66 classified as preferring the more frequently shown bag out of the 97 participants. There was again no difference between the two conditions, $BF_{01} = 6.61$, $p = .377$, but a strong overall deviation from equally distributed groups, $BF_{+0} = 160,646.31$, $p < .001$.

Base rate estimate premeasure. We calculated a log-transformed base rate ratio from these estimates that allowed us to quantify the perceived skewness of participants' estimates. The larger the log score, the more strongly the skewness of the base rates in the same direction, while scores around zero indicate no skew on either variable and negative scores indicate skews in opposite directions. For the log-transformation we used the following formula (Kutzner, 2009):

$$\log_{BR} = \log_{10}\left(\frac{ab}{cd}\right) \times \log_{10}\left(\frac{ac}{bd}\right) \quad (4)$$

with ab, cd, ac, and bd being the base rates of a standard four-cell contingency table (thus, $ab = a + b$, etc.).

It is immediately apparent in the means, that participants perceived the evidence they encountered as quite skewed. They estimated to have encountered the frequent bag 66.51% ($SD = 14.80$) but the infrequent bag only 36.40% ($SD = 13.61$) of the time. Likewise, they estimated to have encountered the frequent outcome 70.24% ($SD = 14.46$) but the infrequent outcome 32.06% ($SD = 15.01$) of the time.⁵ These estimates of the initial evidence are regressive but reasonably accurate. The mean log-score in the reward-rich ($\log_{\text{rich}} = .12$, $SD = 0.16$) and reward-impooverished ($\log_{\text{impooverished}} = .13$, $SD = 0.20$) condition did not differ from one another, $BF_{01} = 6.22$, $t(185.36) = -0.29$, $p = .77$, $d = -0.04$, 95% CI_d [-0.32, 0.24], but differed strongly from zero, $BF_{+0} > 1,000,000.00$, $t(197) = 9.76$, $p < .001$, $d = 0.69$, 95% CI_d [0.54, 0.85], suggesting that participants did indeed perceive strong skews in the same direction for the distribution of both the bags and the balls.

Conditional probability estimate premeasure. Participants estimated that they drew the frequent color 65.42% ($SD = 18.88$) of the time when choosing the frequent bag but only 49.58% ($SD = 21.74$) of the time when choosing the infrequent bag. These estimates closely mirror the estimates participants made in Experiment 2a. The mean ΔP -score in the reward-rich ($\Delta P = .16$, $SD = 0.32$) and reward-impooverished ($\Delta P = .16$, $SD = 0.32$) condition did not differ from one another, $BF_{01} = 6.45$, $t(195.8) = -0.08$, $p = .94$, $d = -0.01$, 95% CI_d [-0.29, 0.27], but again indicated a large overall difference from chance level, $BF_{+0} > 1,000,000.00$, $t(197) = 6.94$, $p < .001$, $d = 0.49$, 95% CI_d [0.35, 0.64].

Confidence premeasure. Participants estimated their confidence similarly in the reward-rich ($\Delta P = .03$, $SD = 0.20$) and the reward-impooverished condition ($\Delta P = .04$, $SD = 0.19$), $BF_{01} = 3.69$, $t(393.47) = 1.36$, $p = .175$, $d = 0.14$, 95% CI_d [-0.06, 0.33].

Sampling. Replicating our previous results, participants again showed biases in sampling in the reward-rich condition but attenuated any biases in the reward-impooverished condition as can be

⁵ We did not force participants to make estimates that would add up perfectly to 100 and therefore report here these uncorrected scores. The results are virtually identical with corrected scores.

seen in Figures 10 and 11. Over the 84 trials of the free sampling phase, participants in the reward-rich condition sampled the frequent bag on average 56% ($SD = 24.36$) of the time. In the reward-impooverished condition on the other hand, participants sampled the frequent bag on average 44% ($SD = 15.21$) of the time. In other words, we found a bias in the reward-rich condition, $BF_{+0} = 8.56$, $t(97) = 2.79$, $p = .003$, $d = 0.28$, 95% CI_d [0.08, 0.48]. However, participants also differed from chance-level in the reward-impooverished condition with the data indicating choices for the frequent option to have been, on average, below chance-level (and hence the test having a negative sign), $BF_{01} = 0.04$, $t(99) = -3.43$, $p < .001$, $d = -0.34$, 95% CI_d [-0.54, -0.14]. The strength of the bias did not differ between the two conditions, $BF_{+0} = 0.26$, $t(162.08) = 0.57$, $p = .284$, $d = 0.08$, 95% CI_d [-0.20, 0.36].

As before, we then analyzed participants' choices over time with a linear mixed model with participants as random effect and trial number, condition, and their interaction as fixed effects. The nonsignificant intercept indicates the quick attenuation to chance level in the reward-impooverished condition ($c = -0.20$, $z = -1.55$, $p = .121$). The main effect for condition indicates a successful bias induction in the reward-rich condition ($\beta_{\text{condition}} = 0.71$, $z = 3.83$, $p < .001$). The negative estimate for trial, however, indicates not attenuation but a strengthening of the bias across trials in the reward-impooverished condition. The nonsignificant interaction term, finally, suggests that the difference between conditions did not change over time ($\beta_{\text{trial} \times \text{condition}} = -0.10$, $z = -0.69$, $p = .488$). Finally, shifting the trial number indicated both a lingering bias in the reward-impooverished condition ($c = -0.32$, $z = 2.23$, $p = .026$) as well as the reward-rich condition ($\beta_{\text{condition}} = 0.62$, $z = 3.16$, $p = .002$).

We again found the predicted effects on oscillation between the two options. Participants in the reward-impooverished condition switched more often ($M = 33.57$, $SD = 5.34$) than participants in

the reward-rich condition ($M = 27.83$, $SD = 5.21$), $BF_{+0} > 1,000,000.00$, $t(165.89) = 7.05$, $p < .001$, $d = 1.09$, 95% CI_d [0.76, 1.41]. We found a pattern very similar to Experiment 2a with 57.0% of participants in the reward-impooverished and 36.7% of participants in the reward-rich condition switching on the first trial, as can be seen in Figure 12.

Relative contingency estimate postmeasure. Following the free sampling phase, we once again expected participants to have maintained their initial bias in the reward-rich, but to have attenuated the bias in the reward-impooverished condition. Indeed as can also be seen in Figure 13, participants exhibited a bias in the reward-rich condition ($M = 7.84$, $SD = 30.54$), $BF_{+0} = 4.65$, $t(97) = 2.54$, $p = .006$, $d = 0.26$, 95% CI_d [0.06, 0.46], but not in the reward-impooverished condition, demonstrated by the expected null effect ($M = 4.29$, $SD = 25.12$), $BF_{01} = 2.23$, $t(99) = 1.71$, $p = .091$, $d = 0.17$, 95% CI_d [-0.03, 0.37]. However, we do not find support for the strength of the effect, $BF_{+0} = 0.36$, $t(187.51) = 0.89$, $p = .187$, $d = 0.13$, 95% CI_d [-0.15, 0.41].

Seven participants were excluded from the binary analyses due to their relative contingency equaling exactly zero. In the reward-rich condition 58 participants preferred the frequent bag out of the 94 participants in this condition, $BF_{+0} = 3.31$, $p = .015$. In the reward-impooverished condition 60 out of the 97 classified as preferring the frequent bag which also differs from chance-level behavior, $BF_{10} = 1.93$, $p = .013$. To test the strength of the effect, we then tested the proportion in the reward-rich against the proportion in the reward-impooverished condition finding no support for a difference between the two conditions, $BF_{+0} = 0.21$, $p = .558$.

Base rate estimate postmeasure. Interestingly, while participants estimated to have sampled both options less skewed (frequent: $M = 53.86$, $SD = 24.60$; infrequent: $M = 49.80$, $SD = 24.04$), they estimated the outcomes they encountered as more skewed than after the initial evidence (frequent: $M = 79.03$, $SD = 9.66$; infrequent: $M = 23.39$, $SD = 11.18$). For

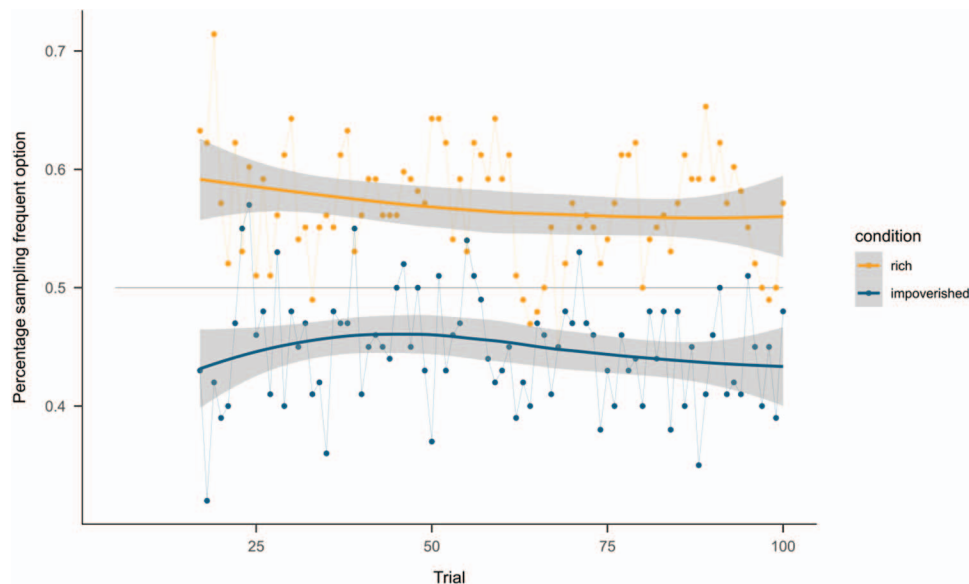


Figure 10. Percentage of participants sampling the frequent option per trial (Experiment 2b). See the online article for the color version of this figure.

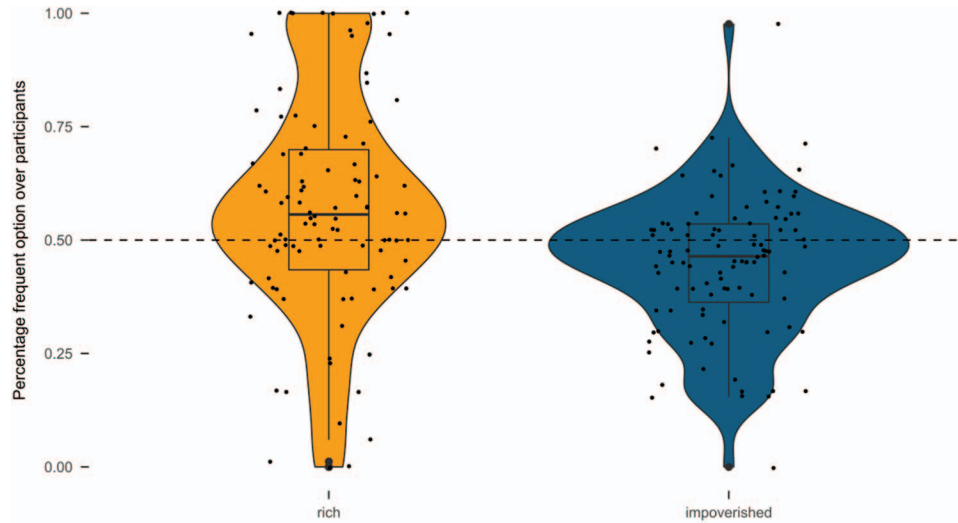


Figure 11. Proportion of choosing the frequent option per participant (Experiment 2b). See the online article for the color version of this figure.

participants in the reward-rich condition, these estimates resulted in a mean log-score of $\log_{\text{rich}} = .13$ ($SD = 0.45$), which is about 12 times more likely to represent the predicted effect of continued perception of skewness in the encountered evidence, $BF_{+0} = 12.78$, $t(97) = 2.95$, $p = .002$, $d = 0.30$, 95% $CI_d [0.10, 0.50]$. Participants' estimates in the reward-impoverished condition resulted in a mean log-score of $\log_{\text{impoverished}} = -.06$ ($SD = 0.30$), with little evidence for either the expected null effect or the alternative hypothesis, $BF_{01} = 1.02$, $t(99) = -2.14$, $p = .035$, $d = -0.21$, 95% $CI_d [-0.41, -0.02]$, but the descriptive suggestion that there might be a tendency of the skewness to be reversed. There was also no difference

between conditions, $BF_{+0} = 0.61$, $t(167.72) = 1.29$, $p = .099$, $d = 0.18$, 95% $CI_d [-0.10, 0.47]$, see also Figure 14. Interestingly, the log-ratio correlates highly with the sampling index, $r = .80$ for both conditions indicating a relationship between the perceived skewness of the base rates and the exhibited sampling behavior (cf. Kutzner, 2009). In other words, the more biased sampling was, the more extremely participants perceived the distributions of options and outcomes to be skewed. The maintenance or attenuation of biases in the reward-rich and reward-impoverished condition, respectively, is therefore also apparent in the perceived skewness participants report after the sampling phase.

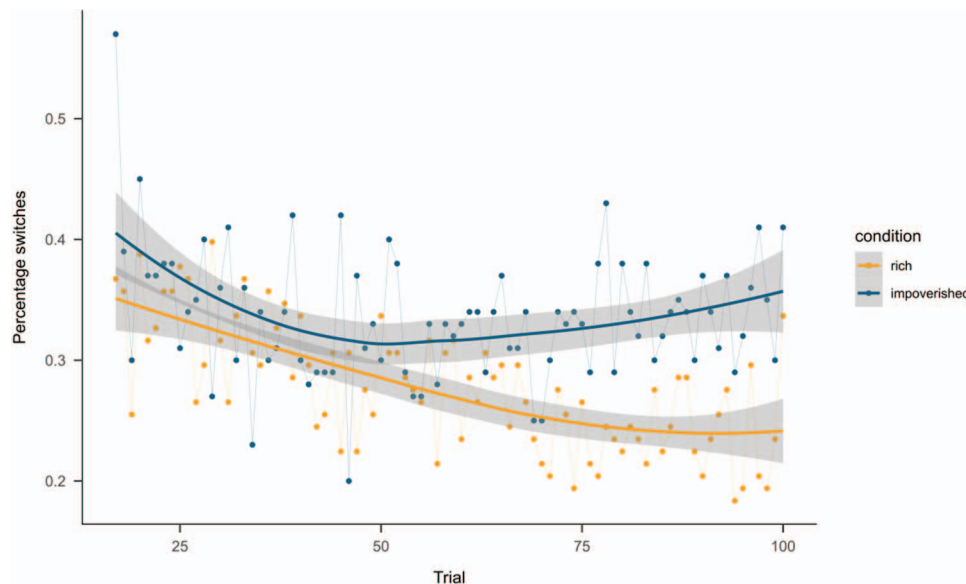


Figure 12. Percentage of participants switching from previous choice including trend lines per condition (Experiment 2b). See the online article for the color version of this figure.

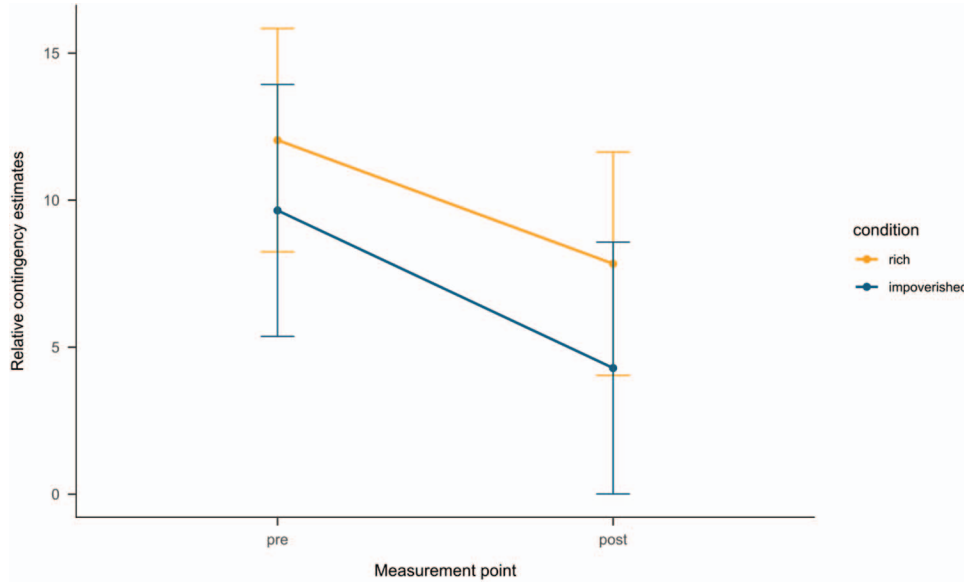


Figure 13. Relative contingency estimates for the pre and post measurement for the frequent over the infrequent bag (Experiment 2b). See the online article for the color version of this figure.

Conditional probability estimate postmeasure. Participants' estimates for grabbing a ball of the frequent (winning) color in the reward-rich condition (frequent bag: 67.14%, $SD = 18.91$, infrequent bag: 59.64%, $SD = 19.80$) resulted in a mean ΔP -score of $\Delta P = .08$ ($SD = 0.30$), which is about two times more likely to represent the predicted effect of a pseudocontingency inference in favor of the frequent option, $BF_{+0} = 4.09$, $t(97) = 2.49$, $p = .007$, $d = 0.25$, 95% $CI_d [0.05, 0.45]$. Participants' estimates in the reward-impoverished condition (frequent bag: 71.05%, $SD = 19.64$, infrequent bag: 69.47%, $SD = 16.56$), on

the other hand, resulted in a mean ΔP -score of $\Delta P = .02$ ($SD = 0.20$), which is more likely to reflect a null effect, $BF_{01} = 6.64$, $t(99) = 0.80$, $p = .427$, $d = 0.08$, 95% $CI_d [-0.12, 0.28]$. There was, however, no support for the strength of the effect, $BF_{+0} = 1.03$, $t(168.13) = 1.64$, $p = .051$, $d = 0.23$, 95% $CI_d [-0.05, 0.51]$. See also Figure 15.

Confidence postmeasure. Finally, participants again did not estimate their confidence in their judgments to be different in the reward-rich ($\Delta P = .02$, $SD = 0.28$) compared with the reward-impoverished ($\Delta P = .01$, $SD = 0.15$) condition, $BF_{01} =$

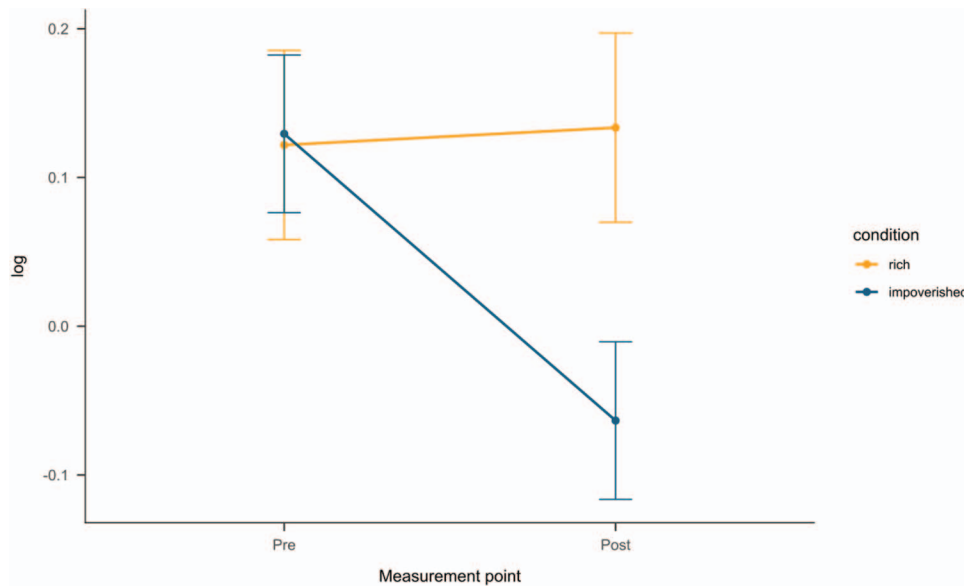


Figure 14. ΔP -scores from base rate estimates (Experiment 2b). See the online article for the color version of this figure.

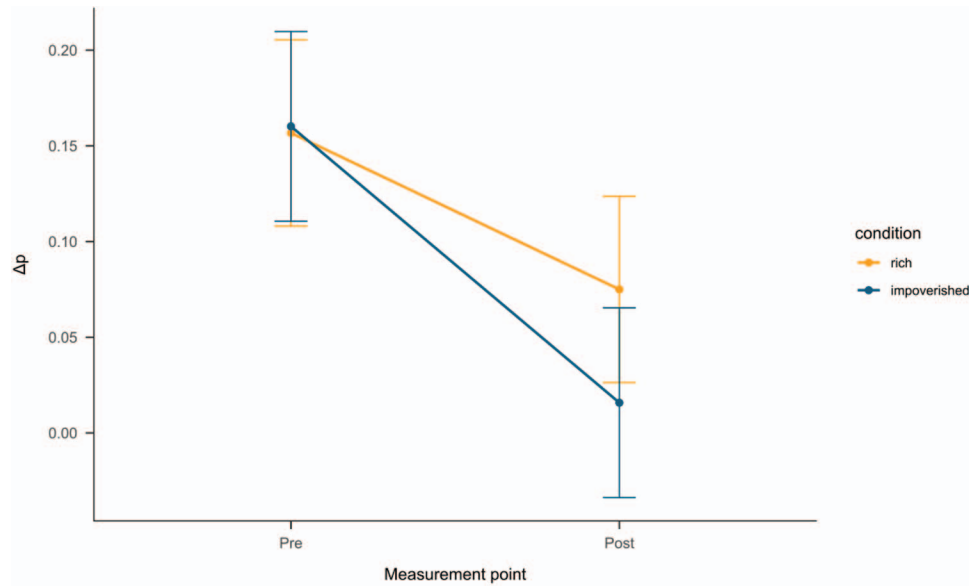


Figure 15. ΔP -scores from conditional estimates (Experiment 2b). See the online article for the color version of this figure.

7.35, $t(392.1) = 0.65$, $p = .518$, $d = 0.06$, 95% CI_d [-0.13, 0.26].

Discussion

In general, this replication produced a pattern very similar to Experiment 2a. However, there are some notable exceptions. Most striking is the maintained bias during sampling in the reward-impoverished condition. Instead of attenuating quickly to chance level, participants maintained a preference for the option that in the initial evidence would seem more favorable (the infrequent option being associated with the infrequent outcome, wins). When it came to making final estimations, however, participants reported a mixed pattern of estimates. Overall, the estimation measures suggest attenuation of any bias. Descriptively, however, the relative contingency measure indicates a reversal of this bias toward the frequent option. While it is very much possible that these results are random fluctuations, they could also be due to the particular circumstances of this lab replication. First, participants did this task as part of a longer session and so fatigue or boredom might play a larger role here relative to Experiments 1 and 2a. Additionally, while on Prolific the study description clearly states the incentivized payoff scheme, this scheme may have been less apparent to participants in the lab replication. And, even more importantly, relative to the overall payoff (due to the duration of the entire procedure) the incentives for this particular study were lower compared with the online studies. That is, the incentives may have been too low or not highlighted enough to truly motivate biased processing in participants. As such, participants may not have felt wins and losses as sensitively as they had in the previous experiments. Nonetheless, the overall emerging pattern is very much in line with our hypotheses and furthermore validates the experimental task as well as the results beyond an online study setting. To summarize then, across the three experiments we find a strong recurring pattern

of persisting biases in the reward-rich condition across both sampling and the later relative contingency and conditional probability estimates. We find a similarly consistent pattern of attenuation of initial biases in the reward-impoverished condition during sampling that also reflects in the estimations thereafter. But we find mixed results when it comes to the strength of the effect.

General Discussion

The present research tested the hypothesis that initial biases are maintained in reward-rich but attenuated in reward-impoverished environments. We first ran simulations of three different learning models that support the above-stated hypothesis for learning models that incorporated base rate sensitivity, but not for models that did not incorporate this sensitivity. In Experiment 1 we found first empirical support for this hypothesis. In Experiment 2a we improved the induction of an initial bias by utilizing a distribution of initial evidence which represented the overall distribution. We generally found support for the successful induction of a bias through the initial evidence as well as strong evidence that biases were maintained in the reward-rich but attenuated in the reward-impoverished condition. These results were backed in a replication study (Experiment 2b). We repeatedly found that frequent rewarding outcomes led to the maintenance of initial biases in an interaction of primacy effects and the rewarding environment. In reward-impoverished environments, on the other hand, these initial biases were attenuated as the frequent negative outcomes discouraged premature exploitation.

It should be noted, however, that the results are not perfectly in line with our predictions. First, and perhaps most importantly, we do see evidence of the bias not always reliably persisting across all measures for the entirety of the experiments. Perhaps this should not be surprising as the mechanism

we describe hinges on exploitation while repeatedly choosing one and the same option is a dull task and the reward perhaps not high enough to commit participants. It is also possible that participants in our experiments are much better at detecting the contingencies than we had hypothesized and attenuation is only slower in the reward-rich condition. While we cannot rule out this possibility, the fact that participants would be so much slower in detecting the underlying contingencies in the reward-rich compared with the reward-impooverished condition in and of itself would also already be fascinating.

Second, testing the difference between conditions in how much they deviate from chance level (what we called the strength of the effect in the above analyses), does not reliably indicate differences between conditions as we had hypothesized. There are two contributing factors, namely the slight attenuation we at times see in the reward-rich condition as well as the variance of the reward-impooverished condition around chance level. Nonetheless, across all measures and experiments, we do find consistent evidence for a difference between the two conditions.

Finally, and perhaps most surprising, are the results of the replication in Experiment 2b in which participants maintained a bias in the reward-impooverished condition during sampling. As discussed above, this is likely to be due to the particular circumstances of the lab replication. But alternative explanations should not be ruled out.

The current findings extend the pseudocontingency literature to the domain of active and repeated sampling. Traditionally, this literature has focused on the formation of contingency inferences but has not made any predictions about choice behavior. To our knowledge, only a single study so far has combined the pseudocontingency literature with choice behavior (Meiser et al., 2018). Over four experiments, the authors repeatedly showed that the skewed distributions elicited pseudocontingency inferences and lead to biased preferences, that is, choices in line with the pseudocontingencies. The first measurement point of Experiment 2a conceptually replicates their findings. In the current studies, however, we focus on how such biases are maintained or attenuated as people continue interacting with their environment and the interaction between initial biases and the reward-structure of said environment comes to play. Our findings show that when a clear-cut reward structure encourages exploration and thereby undermines the skewed distribution of observations about both options (i.e., undoing the premise of pseudocontingency inferences), the initial pseudocontingency effect disappears, reflecting the enduring effect of the reward structure (see also Kareev, Fiedler, & Avrahami, 2009).

What is more, the current findings also extend the large literature on the influence of preferences on sampling (Denrell, 2005; Higgins, 1997) by adding a crucial moderator in the form of the environment decision makers encounter. While reward-rich environments can easily lead to maintained biases even as we continue interacting with our surroundings and try to exploit the positive outcomes the environment yields, reward-impooverished environments are more likely to quickly lead to unbiased mental representations of the environment.

Alternative Explanations

It might seem like the results could be explained as a confirmation bias. Work on this bias has revealed that people are more likely not only to sample but also to integrate information that confirms rather than disconfirms their beliefs (Klayman, 1995; Nickerson, 1998). Crucially, theory on conformation bias assumes

that decision makers exhibit such a bias because they are motivated to uphold a particular belief instead of reaching the most objective conclusion (Klein & Kunda, 1992; Kunda, 1990; Lord, Ross, & Lepper, 1979). However, we argue that the main motivation in the current experiment was to earn as much as possible over the duration of the task and that the exploitation behavior exhibited by participants in the reward-rich condition is due to reward pursuit. Hence, as participants were not motivated to uphold a particular belief, the current findings are not easily explained as a classic conformation bias effect.

It would be possible, that participants did not display a confirmation bias in the traditional, motivational sense but followed a positive test strategy (Klayman & Ha, 1987). It may well be that as long as most outcomes were rewarding (i.e., in the reward-rich but not in the reward-impooverished condition), participants chose the option they expected to confirm rather than disconfirm their beliefs. Such processes are, however, largely associated with information acquisition. That is, as people explore alternatives, they may readily follow such a positive test strategy trying to confirm their initial hunches. On a phenomenon level, positive test strategy may readily lead to behavior identical to that of biased exploitation. The difference is, however, the underlying process that is either one of biased information acquisition (positive test strategy) or reward pursuit (biased exploitation). Given the extensive sampling phase of over 80 trials and the attenuation of biases in the reward-impooverished condition, biased exploitation seems a more likely explanation than a positive test strategy, but we cannot rule out the latter completely.

It should come as no surprise that many reinforcement learning models (e.g., the Rescorla-Wagner model; Rescorla & Wagner, 1972) as well as Bayesian updating accounts (e.g., Körding & Wolpert, 2004; Yu, 2007) would make predictions that are at least partially in line with the claims we have laid out here, as those literatures are well developed and can generally model learning and belief-updating well. Nonetheless, there are some specific distinctions from our account which are worthwhile highlighting. The results of Experiment 1 might still readily be explained in terms of reinforcement learning accounts. In the reward-rich condition the actual contingency of initial evidence would have led participants to develop strong prior beliefs. Further sampling reinforced these beliefs leading to the maintenance of the initial bias. In the reward-impooverished condition, on the other hand, positive reinforcement was rare and, accordingly, attenuation of the initial bias was likely. Note, however, that our simulations using the Rescorla-Wagner model showed quick attenuation toward chance level and no differences between conditions.

Many of these learning models would usually not predict contingencies to be learned from the initial evidence in Experiments 2a and b. Nonetheless, as one option was more frequent than the alternative, the models might still have sampled and preferred the option on which they had more observations. That is, in the absence of a strong prior belief, confidence may have guided their behavior and estimates (Einhorn & Hogarth, 1978). In the reward-rich condition, we would once again have expected participants to associate the frequent option more strongly with the frequent outcome than the infrequent option. However, we would then also have expected them to estimate their confidence to be higher for the frequent than the infrequent option. In the reward-impooverished condition, we would have also expected participants

to associate the frequent option more strongly with the frequent outcome than the infrequent option and have higher confidence in the frequent option initially, that is before the reward-scheme was known. After the reward-scheme became known, however, we would expect them to quickly lose this bias as the more frequent option was now thought to be the worse option.

Instead, across all experiments and all measurement points do we consistently find that participants' confidence estimates do not differ for the frequent and infrequent bag. Following the initial evidence, they repeatedly align the frequent action with the frequent outcome and just as confidently align the infrequent action with the infrequent outcome as predicted by a pseudocontingency account (Fiedler et al., 2009). Likewise, following sampling of both options they still were equally confident in estimating both alternatives. We predict therefore that in reward-rich conditions any initial bias⁶ can lead to exploitation. Exploitation implies biased sampling and this in turn leads to biased inferences as the skewed distribution is maintained which in turn invites more biased sampling. In reward-impooverished conditions on the other hand, any initial bias will be overcome as participants have no incentive to settle on any particular option and engage in explorative sampling behavior.

That is not to say, that learning models cannot also explain the pattern we simulate and empirically test. As our simulations show, learning models that do not rely solely on an updating rule but instead are sensitive to the underlying base rates, also readily predict the pattern we find. The important distinction between classic reinforcement learning models, such as the Rescorla-Wagner model, and, for example, BIAS or MDM is their reliance on either a single value which is updated repeatedly, or exemplar-based memory representations (though these representations can readily be constructed or aggregated representations) that allow for base rate representations. In the latter instance, learning models can describe the same phenomenon we tackle from an exploration/exploitation perspective.

Comparing the results from our simulations with participants' behavior on the experiments suggests that our participants did indeed incorporate base rate information as the Rescorla-Wagner model offered the worst description of participants' behavior and BIAS offered the best description. Across all models we find initial biases that speak to the strong influence small samples can have on decision making. In the BIAS model in particular, but also for our Bayesian model, we see attenuation of initial biases toward chance level in the reward-impooverished condition. In general, we see less attenuation and instead the persistence of initial biases in the reward-rich condition for both BIAS and the Bayesian model. Participants usually did attenuate more strongly toward chance level than the learning models in Experiments 1 and 2a, but not so in Experiment 2b. Participants in the reward-rich condition attenuated more strongly than the simulated learning models but also maintained initial biases. While the patterns of the simulated models and participants' behavior are not identical and while we did not formally fit any of the models to participants' behavior, this comparison sheds light on the mechanisms at hand: Sensitivity to one's sampling history and initial evidence which results in premature exploitation can lead to persisting biases even in repeated sampling situations. The implication hereof is that while it may for the most part be highly adaptive for a decision maker to remember

their sampling history, in some situations this can lead to unwarranted biases and conclusions about one's environment.

Outlook

To return to the initial question posed, why do humans then develop strong and persisting, but erroneous beliefs? By exploiting options in order to maximize their rewards in the here and now, decision makers dial in on the informational input they receive from this option. Exploitation, by definition, implies sampling certain options more than others and the chances that primacy effects are adjusted sufficiently are accordingly low.

These inferences may explain why people develop unwarranted beliefs about actions and outcomes, such as a belief regarding the administration of alternative medicines. Many of the maladies that humans in Western societies encounter more regularly and that we would be more likely to treat without the consultation of professionals, include maladies such as the flu or common cold. While these obviously have a negative impact, they could, compared with other medical conditions, nonetheless be considered reward-rich environments in that symptoms are usually easily treatable, and we recover within a short period of time. That is, even without medical treatment a relatively quick full recovery, the outcome we all seek, is the most likely outcome (compared with, e.g., a longer, more cumbersome recovery or even no recovery). Framed as such, the process can be explained in the terms of this research: People with prior beliefs regarding the effectiveness of a particular treatment exploit this treatment whenever they have the flu. There is little interest for most of us in exploring different treatments, a quick recovery is paramount. And as this is the most likely outcome anyway, we end up building a distribution of evidence that is strongly skewed toward the action of using a particular treatment and a quick recovery as the most frequent outcome.

Medical examples are also helpful in pointing out the potential consequences of maintaining unwarranted beliefs. While in the experimental procedure above maintaining a bias did no real harm, there can also be serious costs involved. For society, that has to cover treatment costs through insurance companies, and for individuals that spend time, money, and potentially their health on suboptimal treatments.

Environments may also be a strong influencing factor when it comes to maintaining or attenuating first impressions. In his inspiring work, Denrell (2005) argued that the pursuit of positive interactions alone can explain a negativity bias toward others, thereby highlighting the importance of first impressions from a cognitive ecological perspective (Fiedler & Wänke, 2009). He argued that repeated interactions (e.g., due to proximity) can help overcome such initial biases. We would like to add that overcoming initial biases depends not only on repeated interactions alone, but also the environment we find ourselves in: If the environment is reward-rich in that most people do try to make good impressions, we should be more likely to uphold our initial biases. Only

⁶ Here, we use distributions with actual contingencies (Experiment 1) and skewed base rates likely to induce pseudocontingencies (Experiment 2a and 2b). But it would be easily conceivable how, for example, prior beliefs (e.g. communicated beliefs; Pilditch & Custers, 2018; Pilditch et al., 2020) or random fluctuations in the environment might also induce biases.

in environments in which few people are nice toward us would we be likely to readily overcome these initial biases.

Humans and all agentic organisms face the inherent trade-off between information search and reward maximization. Do we gather more information in the hopes of making better decisions or do we continue sampling as many of the rewarding outcomes as possible given the knowledge we currently have? The environments we find ourselves in may heavily shift this trade-off and as a consequence either lead to the maintenance or attenuation of biases. If we do not want to fall prey to initial biases, we might be strongly advised to consider alternatives every now and then—especially if things are going well.

Context of the Research

The importance of first impressions is widely accepted, also in lay psychology. But how do first impressions arise, how do they influence subsequent cognition and behavior, and under what circumstances do these influences persist and when are they attenuated? Our research group with members from the University of Heidelberg and Utrecht University investigates how initial beliefs are updated during continued interaction with the environment. Building on previous work by Pilditch and Custers (2018), this is the first article in a new series that will result in a dissertation. Here we lay out and test our theory of bias maintenance or attenuation as an interaction between sampling behavior and environmental constraints. In later articles we aim to generalize these findings across contexts.

References

- Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*, 147–149. <http://dx.doi.org/10.3758/BF03334492>
- Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology, 2*, 1–9. <http://dx.doi.org/10.1037/h0021966>
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290. <http://dx.doi.org/10.1037/h0055756>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 362*, 933–942. <http://dx.doi.org/10.1098/rstb.2007.2098>
- DeCoster, J., & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review, 8*, 2–27. http://dx.doi.org/10.1207/S15327957PSPR0801_1
- Dennis, M. J., & Ahn, W.-K. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition, 29*, 152–164. <http://dx.doi.org/10.3758/BF03195749>
- Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112*, 951–978. <http://dx.doi.org/10.1037/0033-295X.112.4.951>
- Denrell, J., & Le Mens, G. (2012). Social judgments from adaptive samples. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 151–169). New York, NY: Psychology Press.
- Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*, 180–209. <http://dx.doi.org/10.1037/0033-295X.106.1.180>
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*, 395–416. <http://dx.doi.org/10.1037/0033-295X.85.5.395>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. <http://dx.doi.org/10.3758/BF03193146>
- Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87*, 293–311. <http://dx.doi.org/10.1037/0022-3514.87.3.293>
- Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*, 193–214. <http://dx.doi.org/10.1037/0033-295X.103.1.193>
- Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review, 17*, 504–509. <http://dx.doi.org/10.3758/PBR.17.4.504>
- Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology, 87*, 453–467. <http://dx.doi.org/10.1037/0022-3514.87.4.453>
- Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*, 187–206. <http://dx.doi.org/10.1037/a0014480>
- Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically Unwarranted but Smart Inferences. *Current Directions in Psychological Science, 22*, 324–329. <http://dx.doi.org/10.1177/0963721413480171>
- Fiedler, K., & Wänke, M. (2009). The cognitive-ecological approach to rationality in social psychology. *Social Cognition, 27*, 699–732. <http://dx.doi.org/10.1521/soco.2009.27.5.699>
- Harris, C. (2020). *Repository for: Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments*. Retrieved from <https://osf.io/j3ve4/>
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*, 1280–1300. <http://dx.doi.org/10.1037/0003-066X.52.12.1280>
- Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences, 19*, 46–54. <http://dx.doi.org/10.1016/j.tics.2014.10.004>
- Hojitink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods, 24*, 539–556. <http://dx.doi.org/10.1037/met0000201>
- Jones, E. E., Goethals, G. R., Kennington, G. E., & Severance, L. J. (1972). Primacy and assimilation in the attribution process: The stable entity proposition. *Journal of Personality, 40*, 250–274. <http://dx.doi.org/10.1111/j.1467-6494.1972.tb01002.x>
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263–269. [http://dx.doi.org/10.1016/0010-0277\(95\)92814-G](http://dx.doi.org/10.1016/0010-0277(95)92814-G)
- Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*, 397–402. <http://dx.doi.org/10.1037/0033-295X.107.2.397>
- Kareev, Y., Fiedler, K., & Avrahami, J. (2009). Base rates, contingencies, and prediction behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 371–380. <http://dx.doi.org/10.1037/a0014545>
- Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental*

- Psychology: General*, 126, 278–287. <http://dx.doi.org/10.1037/0096-3445.126.3.278>
- Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 32, pp. 385–418). New York, NY: Academic Press.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228. <http://dx.doi.org/10.1037/0033-295X.94.2.211>
- Klein, W. M., & Kunda, Z. (1992). Motivated person perception: Constructing justifications for desired beliefs. *Journal of Experimental Social Psychology*, 28, 145–168. [http://dx.doi.org/10.1016/0022-1031\(92\)90036-J](http://dx.doi.org/10.1016/0022-1031(92)90036-J)
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427, 244–247. <http://dx.doi.org/10.1038/nature02169>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498. <http://dx.doi.org/10.1037/0033-2909.108.3.480>
- Kutzner, F. (2009). *Pseudocontingencies - rule based and associative*. (Doctoral thesis). University of Heidelberg, Heidelberg, Germany. Retrieved from http://archiv.ub.uni-heidelberg.de/volltextserver/9912/1/Dissertation_Kutzner_UB.pdf
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <http://dx.doi.org/10.18637/jss.v082.i13>
- Lave, C. A., & March, J. G. (1993). *An introduction to models in the social sciences*. Lanham, MD: University Press of America.
- Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. Retrieved from <https://www.sosicisurvey.de>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098–2109. <http://dx.doi.org/10.1037/0022-3514.37.11.2098>
- 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, 191–215. <http://dx.doi.org/10.1037/dec0000033>
- Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, 36, 315–336. <http://dx.doi.org/10.1002/ejsp.304>
- Meiser, T., Rummel, J., & Fleig, H. (2018). Pseudocontingencies and choice behavior in probabilistic environments with context-dependent outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 50–67. <http://dx.doi.org/10.1037/xlm0000432>
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.12–4.2) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220. <http://dx.doi.org/10.1037/1089-2680.2.2.175>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14, 1105–1107. <http://dx.doi.org/10.1038/nn.2886>
- Pilditch, T. D., & Custers, R. (2018). Communicated beliefs about action-outcomes: The role of initial confirmation in the adoption and maintenance of unsupported beliefs. *Acta Psychologica*, 184, 46–63. <http://dx.doi.org/10.1016/j.actpsy.2017.04.006>
- Pilditch, T. D., Madsen, J. K., & Custers, R. (2020). False prophets and Cassandra’s curse: The role of credibility in belief updating. *Acta Psychologica*, 202, 102956. <http://dx.doi.org/10.1016/j.actpsy.2019.102956>
- Prager, J., Krueger, J. I., & Fiedler, K. (2018). Towards a deeper understanding of impression formation—New insights gained from a cognitive-ecological perspective. *Journal of Personality and Social Psychology*, 115, 379–397. <http://dx.doi.org/10.1037/pspa0000123>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rescorla, R., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. <http://dx.doi.org/10.2139/ssrn.2160588>
- Skinner, B. F. (1948). Superstition in the pigeon. *Journal of Experimental Psychology*, 38, 168–172. <http://dx.doi.org/10.1037/h0055873>
- Staudinger, M. R., & Büchel, C. (2013). How initial confirmatory experience potentiates the detrimental influence of bad advice. *NeuroImage*, 76, 125–133. <http://dx.doi.org/10.1016/j.neuroimage.2013.02.074>
- Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, 39, 212–222. <http://dx.doi.org/10.2307/1415413>
- Vulkan, N. (2000). An economist’s perspective on probability matching. *Journal of Economic Surveys*, 14, 101–118. <http://dx.doi.org/10.1111/1467-6419.00106>
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A grammar of data manipulation (R package version 0.8.0.1) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Yu, A. J. (2007). Adaptive behavior: Humans act as Bayesian learners. *Current Biology*, 17, R977–R980. <http://dx.doi.org/10.1016/j.cub.2007.09.007>

Received April 23, 2019

Revision received January 30, 2020

Accepted February 2, 2020 ■