# What You Get Is What You See:

## How the Evidence We Obtain Through Our Behavior Shapes Our Perception of the World

**Chris Harris**

# What You Get Is What You See: How the Evidence We Obtain Through Our Behavior Shapes Our Perception of the World

Chris Harris

**Colophon**
ISBN: 978-94-6423-734-4
Printing: www.proefschriftmaken.nl
Illustration: Maartje Schalker

# What You Get Is What You See: How the Evidence We Obtain Through Our Behavior Shapes Our Perception of the World

**Wat je krijgt is wat je ziet: Hoe de informatie die we door onze keuzes krijgen onze waarneming van de wereld vormt**

**(met een samenvatting in het Nederlands)**

**Proefschrift**

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 8 april 2022 des ochtends te 10.15 uur

door

**Christoph Alexander Harris**

geboren op 16 januari 1990
te Malsch, Duitsland

**Promotoren:**

Prof. dr. H. Aarts

Prof. dr. K. Fiedler

**Copromotor:**

Dr. R. Custers

# Table of Contents

# CHAPTER 1

# Chapter 1:
## Introduction and Overview

Imagine going on a first date. If the date goes well there is a good chance of follow-up dates which may or may not result in a relationship. If the first date does not go well, however, the probability that you go on a second date drops drastically. This might have been the right decision to make. But it might be that because of one negative experience you never give yourself the chance to obtain more information about the other person and update your initial first impression when there could have been the chance of a shared future.

This example, adapted from Denrell (2005), highlights how our behavior influences what information and feedback we might possibly encounter as we interact with the world. If we go on further dates, we can learn more about this person. But by deciding to not go on any further dates, we deny ourselves the opportunity to update our impression about this person. This deceptively simple relationship between our behavior, the evidence we encounter, and how we can then update our beliefs lies at the core of this dissertation. In the following chapters, I will present my empirical research on this interaction and give my perspective of how my work can be placed within the larger context of the field. I will address whether and how, despite plenty of opportunities to put one's beliefs to the test, this interaction between behavior, evidence, and beliefs can lead to persisting biases. And I will end by arguing that ultimately, people shape their own idiosyncratic experience of the world.

**Belief-Updating**

At the outset of this project stood the question why people would maintain erroneous beliefs even when they have ample opportunities to test the validity of their beliefs and even when they should supposedly be motivated to learn as well as they could. Why would people believe so strongly in certain alternative medicines when evidence from the medical sciences is thin at best (Barnes et al., 2008; Beer et al., 2016; Ernst, 2010; Shang et al., 2005)? Surely, people's main motivation should be to get healthy as quickly as possible. Moreover, over time, they find themselves repeatedly in the situation where they can put their belief to the test and learn about the effectiveness of their treatments of choice. After all, most of us get sick at least once a year. And yet, alternative medicines remain widely popular. Similarly, why do first impressions have such a lasting influence on people's perceptions of others? Why do people maintain stereotypes despite repeated interaction? Why do we not always attenuate superstitions? In all of these examples, there typically should be plenty of opportunities to put one's beliefs to the test in further interactions – and yet erroneous beliefs often enough persist. I will return to such examples throughout my dissertation as I develop and empirically test a framework that offers an explanation of these intriguing dynamics.

***First Impressions***

Beliefs are bound to start with some first impression or a first bit of information. And well-established literatures on judgement and decision-making point to the many reasons why such first impressions are actually quite difficult to get right: Random fluctuations in the environment might suggest a given choice alternative to be better than it actually is. In other words, randomness can lead to streaks that make one option appear to be particularly good. In fact, it is a statistical property of our world that small samples tend to amplify the true underlying correlation (Kareev, 1995, 2000; Kareev et al., 2014; Kareev et al., 1997). Such statistical properties would, of course, be corrected as one continues interacting with one's environment thereby increasing the initial sample. But humans tend to attribute too much importance to small samples of initial evidence (Tversky & Kahneman, 1971), thereby overweighting it in their decision processes (primacy effects; Anderson, 1965; Asch, 1946). Moreover, humans are known to form inferences that might differ from the genuine contingency when evidence is skewed, that is, when one option is encountered or chosen more frequently or one outcome occurs more frequently than others (pseudocontingencies; Fiedler et al., 2009) – a cognitive illusion I will return to throughout this dissertation. Finally, humans are known to take the evidence they encounter at face value, basing their decisions on the probabilities at hand while neglecting the distribution from which these probabilities stem (cognitive myopia; Fiedler, 2012). In conclusion, small samples, on which first impressions are typically based, can easily be misleading. But surely people would overcome these as they continue interacting with their environment and gathering further information?

***Learning vs. Reward Pursuit***

The typical assumption, reminiscent of Bernoulli's (1713) law of large numbers, would be that any initial biases are washed out and attenuated as people continue interacting with their environment, learning new information, and updating their beliefs (Behrens et al., 2007; Harris et al., 2021; Rescorla & Wagner, 1972; Yu, 2007). This would indeed be the case, if people would receive a representative sample of the available options for behavior. However, it almost goes without saying that our beliefs influence our behavior (Ajzen, 1991; Edwards, 1954; Savage, 1954; Tolman, 1934). And they do so whether our beliefs are accurate or not (cf. Bott & Meiser, 2020; Meiser et al., 2018). The question then arises whether people's behavior truly allows them to correctly update their initial beliefs (see also Fiedler, 2008).

If people are invested in learning, they might try out alternatives, explore available options, and could correctly update their beliefs sooner or later (that is, as long as they don't engage in a positive testing strategy; Klayman & Ha, 1987). But

a point that I will be making throughout this dissertation is that learning is typically not the foremost goal. Instead, behavior is often driven by reward pursuit such as, for example, earning as much money as possible during an experiment, having a great date, or enjoying a great dinner at a restaurant. Of course, choosing a good restaurant for dinner requires us to have learned about different restaurants in the neighborhood. But once we have an initial idea, we are typically not interested in trying out restaurants that we expect to be disappointing simply for the sake of learning, just as we are not interested in what we expect to be bad dates simply to get to know another person better. Instead, our choices are driven by higher-level goals, and learning takes place almost incidentally. The important implication, then, is that learning is contingent on our behavior. In order to update our first-hand experiences, we need to encounter new evidence – and whether we do so depends on the choices we make (Denrell, 2005, 2007; Denrell & Le Mens, 2011, 2012; Denrell & March, 2001).

***Exploitation and Information Restriction***

This tradeoff between information search on the one hand and goal pursuit (e.g., reward-maximization) on the other hand, is oftentimes described in terms of exploration and exploitation (Cohen et al., 2007; Hills et al., 2015; Mehlhorn et al., 2015). While exploration describes behavior focused on learning about choice alternatives, exploitation describes the focus on one supposedly best option in order to maximize rewards. Typically, humans do not engage in the most extreme form of exploitation, but instead balance their behavior by occasionally sampling other options (probability matching; Gaissmaier & Schooler, 2008; Vulkan, 2000). But even so, the supposedly best option would still be chosen considerably more frequently than alternative options with the logical consequence that one would receive more information about the frequently chosen option compared to others.

Throughout this dissertation I will argue that this tradeoff, between learning on the one hand and maximizing sought outcomes on the other hand, affects our choices, the evidence we encounter, and our beliefs. I argue that general features of the environment, and more specifically an environment's general abundance or scarcity of positive outcomes, should have a strong influence on whether erroneous beliefs are maintained or attenuated over time. That is, our behavior, and specifically exploitation of a supposedly best option, can result in a sample of evidence that makes the attenuation of any initial bias extremely difficult and that can lead to biases being maintained.

Building on previous work on biases in judgment and decision making, I will propose a more comprehensive framework to address how people experience a subjective version of the world through their choices and the evidence they

encounter. It is this interplay between beliefs, behavior, and the environment that can explain persisting biases and offers an answer to the question posed at the outset of this dissertation project of why people would maintain erroneous beliefs despite repeated opportunities to put said beliefs to the test: Even though there are opportunities in which one can learn about alternative options, people do not make sufficient use of such opportunities because the ulterior goal is the maximization of rewards and not optimal learning. Ironically, then, it is exactly this focus on goal pursuit that can lead to persisting biases, as the motivation lies more so on exploiting a (supposedly) best option than on learning about all alternatives (for an overview see also Harris & Custers, 2022).

To end on a positive note, I would like to emphasize that the conclusion should not be that humans are terrible decision-makers. On the contrary, under many conditions and circumstances we are really quite sensitive to our environment. However, as we have limited resources, hard choices have to be made. To return to the incipient example of going on a date, most people are limited in the amount of money and energy they can invest in repeated dates, and no one has unlimited time. As a result, choices have to be made in which less promising dates are dropped in favor of the more promising ones. It is a truism of the world we live in that we cannot always engage in sufficient information search to learn the true underlying probabilities of all choice alternatives (Brunswik, 1952; Payne et al., 1993). And so, if a decision maker believes to have found a best alternative it can make perfect sense that they begin concentrating their efforts on this option and neglect other, presumably worse, options. The caveat, of course, is that under certain circumstances this can lead to persisting biases as described in this dissertation. This does not mean that such biases are not reversible or avoidable. At work or via common friends, we are forced to repeatedly interact with people even if we initially did not like them. This, in turn, forces us to learn more information about them and we may update negative initial impressions. Similarly, we may read or be told about other restaurants, thereby learning new information without ever having gone there ourselves. And, of course, better understanding these processes also offers opportunities for directed interventions. Nonetheless, the ultimate test for one's beliefs remain the firsthand experiences we make, and in this dissertation I examine how the interplay between our beliefs, our behavior, and the environment can lead to persisting biases despite best intentions. Examining this interplay is important and crucial to understand and appreciate how we shape the subjective world we live in.

**Outline of the Dissertation**

In the remaining chapters, I differentiate between reward-rich environments, in which most outcomes are positive, and reward-impoverished environments, in

which most outcomes are negative. I explore how the temptation to exploit can lead to persisting biases as a function of the environment one is in. If a person has an initial belief that one option might be better than another option, they might transition their behavior towards exploiting this seemingly best option. And, in a reward-rich environment, in which most outcomes are positive, it would seem that they are being confirmed in their strategy. However, if their initial belief was unwarranted, they might never explore alternative options sufficiently well to notice that other options would have been equally good or potentially even better. After all, why change a (supposedly) winning horse? In reward-impoverished environments, on the other hand, in which most outcomes are negative, a person might as well still have an initial belief about one option being better. However, exploitation of this option would result in frequent losses, making continued exploitation less likely. Instead, the most likely behavior is switching between choice alternatives resulting in exploration that should result in finding the actually best option.

The tasks I use throughout my theses are referred to as two-armed bandit tasks in reference to the slot machines colloquially called bandits. Just like one could play different arms of such a bandit, a two-armed bandit task specifies that a participant will have two choice alternatives from which they can choose repeatedly. For example, in the first paradigm I developed, participants could choose between two bags that contain a certain distribution of yellow and blue balls.

Such tasks are often used in the decision-making literature because they allow the experimenter to track participants' choices over time in a controlled decision environment (Abbott & Sherratt, 2011; Hotaling et al., 2020; Newell et al., 2015; Speekenbrink & Konstantinidis, 2015; Steyvers et al., 2009). Additionally, with such tasks it is easy to add or remove choice alternatives. For example, in the first paradigm selecting one of two bags would result in the participants "grabbing" a yellow or a blue ball. But on the first few trials, I would only present one of the two choice alternatives and control for the outcome. This allowed me to enforce that all participants encountered the same distribution of initial evidence meant to induce a bias, that is, an unwarranted preference towards one of the two alternatives. In later trials, participants were free to choose between the two alternatives.

In many of my experiments I rely on pseudocontingencies for the induction of initial biases, a literature that traditionally was not concerned with actual choice behavior and instead relied on estimation and prediction tasks, often based on evidence that was selected for, and not by, participants (Fiedler & Freytag, 2004; Fiedler et al., 2009; Fiedler et al., 2013). But in my tasks, I only enforced a sample of initial evidence and then observe participants' free choices (and as such their preferences) and how their (pseudocontingency) biases develop over time.

Throughout my thesis, I introduce three critical variations of my two-armed bandit task: Often participants would first encounter a distribution of initial outcomes before learning which outcomes were rewarding. When participants did not yet know the value of outcomes, they had to pay equal attention to all outcome information. However, in everyday life, we often know the value of outcomes. Therefore, in some experiments the value of the outcomes (e.g., positive vs. negative) was known from the start of the experiment on, which might be considered a more realistic task setting. Often outcomes were binary (win or loss), but sometimes they were graded (smaller or larger wins and losses). While binary outcomes simplified the task, graded outcomes might, again, be considered a more realistic task setting but they can also make the relationships more difficult to detect. And often both choice alternatives were identical, but in some experiments one alternative was actually better than the other. In the latter task setting there is a clear better option, which emphasizes that participants are not just lazy but truly have persisting biases. Note, however, that the basic setup – repeated choices between two choice alternatives – always remained the same.

Chapters 2-5 contain my empirical work and are therefore written in the form of empirical papers. As a consequence, some of the information in the chapters will overlap. That being said, Chapter 2 provides a thorough introduction into the theory and methods I rely on throughout this dissertation. Chapter 6 offers a comprehensive discussion and perspective on my findings in the larger context of the field.

In Chapter 2, I address what influence the environmental characteristics (in the form of high or low outcome probabilities) have on the maintenance or attenuation of biases. Specifically, I test the prediction that an initial bias favoring one of two 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. I 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-impoverished 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 non-preferred option. In three 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 one 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. I argue that biased sampling in reward-rich environments could be responsible for false beliefs about the outcomes of behavioral options.

In the two chapters that follow, I test boundary conditions of this concept. As outcomes in real life are rarely binary and not always subject to self-consistency commitment, in Chapter 3, I investigate whether my previous findings still hold true under these new task settings. To address this question, I develop a new experimental task that introduces a more meaningful everyday environment and, importantly, allows for graded outcomes. In three experiments using a modified paradigm with two options leading to graded outcomes, participants were induced to infer a (illusory) contingency favoring one option, before sampling freely between both options. In line with earlier research, the induced initial preferences were either maintained or corrected, depending on participants' sampling strategies. These findings testify to the generality of the phenomenon, which is not peculiar to the restrictive task settings of my research presented in Chapter 2. They increase our understanding of the conditions under which people maintain unwarranted preferences in continued interaction with the environment despite the availability of counterevidence.

In Chapter 4, I investigate the type of bias that is induced as well as the downstream consequences thereof with a modified version of the experimental task from Chapter 3. Specifically, I investigate an idea raised in the earlier chapters that initial knowledge of the valence of all outcomes may be an important moderator for which type of bias participants would form initially. If participants are initially unaware of the valence of outcomes, I expect pseudocontingencies to arise which would then be attenuated as participants continue sampling. If, however, participants are initially aware of the valence of the outcomes, participants may neglect the total number of choices made and instead focus solely on the number of positive outcomes encountered. This denominator neglect (Reyna & Brainerd, 2008) should then result in the persistence of initial biases. I manipulate this potential moderator and find that knowledge over the valence does not seem to be a key moderator in explaining when participants form what type of cognitive bias. The search for the differentiating moderator underlying these biases remains open and intriguing.

In Chapter 5, I return to the initial paradigm of Chapter 2 and address how initial biases can lead to persisting biases even when an alternative is objectively better. In two studies, I demonstrate that in a task in which participants could repeatedly choose between two options to gain rewards, erroneous initial impressions about yielded outcomes can lead to persisting biases towards a clearly inferior option. I again argue that a premature focus on reward pursuit (exploitation) rather than exploration is the cause of these biases, which persist despite plenty

of opportunities and a presumed motivation to overcome them. By focusing on a supposedly best option, participants never give themselves the chance to sufficiently try out alternatives and thereby overcome their initial biases. I conclude that going for the money is not always the best strategy.

In Chapter 6, I present a perspectives paper in which I propose an iterative cycle of decision-making and belief-updating. Once again, this chapter departs from the notion that, in order to realize desirable and rewarding outcomes, or goals, people have to constantly decide what choice alternatives to engage with. I identify three stages - encountering evidence, updating one's beliefs, and making the next decision - that together form this iterative cycle. It is this cycle that often lies at the hearth of persisting biases. For example, prejudices influence whether people interact with others. But if they do not interact with certain people because of negative beliefs they hold, they cannot learn new information about these people and therefore cannot correct their prejudices. Similarly, habitual behavior results in automatic choices that result in encountering information and updating one's beliefs about this option but not others. And finally, as a result of the social bubbles we operate in, we are more likely to encounter evidence that is in favor with the view shared within the social bubble than information that opposes its views. This, however, will once again restrict the extent to which one can possibly update one's beliefs and behavior. The interdependency of the three stages, and the iterative cycle within which I argue that they must be seen, serves to explain the persistence of biases in stereotypes, superstitions, or beliefs in alternative medicine. I argue that this cycle is ubiquitous in psychology and discuss a range of phenomena departing from this perspective.

# CHAPTER 2

# Chapter 2:
## Biased Preferences Through Exploitation: How Initial Biases Are Consolidated in Reward-Rich Environments

**This chapter is based on:**

Harris, C., Fiedler, K., Marien, H., & Custers, R. (2020). Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments. *Journal of Experimental Psychology: General, 149*(10), 1855–1877.

## Abstract

In the current paper, we test the prediction that an initial bias favoring one of two 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-impoverished 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 non-preferred option. In three 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 one 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.

A fundamental tradeoff 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 et al., 2007; Hills et al., 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 paper, 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 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-impoverished 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-impoverished 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-impoverished 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 et al., 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 et al., 1972) in their learning as well as their behavior (Decoster & Claypool, 2004; Pilditch & Custers, 2018; Pilditch et al., 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 et al., 2009; Meiser et al., 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 et al., 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 et al., 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-Impoverished Environments***

An appropriate experimental manipulation of hedonic influences on belief maintenance is to contrast reward-rich and reward-impoverished 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 an erroneous contingency is less likely to be corrected than in a reward-impoverished environment with outcomes being more often negative and less often positive. In a reward-impoverished 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 towards the other option. However, as further

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.

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-impoverished 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 tradeoff 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-impoverished environment would not, crucially shifting the tradeoff 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 et al., 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 paper, we test our claim that erroneous contingency inferences are formed and maintained in reward-rich but attenuated in reward-impoverished 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 one-hundred 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 seventy-five winning and twenty-five losing balls, and a reward-impoverished condition, in which each bag contained seventy-five losing and twenty-five winning balls.

The one-hundred 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 sixteen trials that contained a zero contingency between bags and ball color, but skewed base rates yielded a distinct pseudocontingency. We used these same two distributions of initial evidence (Table 2.1) 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 one hundredth 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.

**Methods**

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 sixteen trials also later utilized in Experiments 1 and 2 (summarized in Table 2.1). 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 b, 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 $n_{col}$ = 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 α = .25, β = 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 $V_{s,t0}$ = 0.5, we used the following updating equation (2.1):

$$V_{s,t} = V_{s,t-1} + \alpha\left(r_{t-1} - V_{s,t-1}\right) \tag{2.1}$$

Along with the following observation equation (2.2):

$$p(s) = \frac{\exp\,(\beta \cdot V_s}{\sum_i \exp(\beta \cdot V_i)} \tag{2.2}$$

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-impoverished 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-impoverished condition (see also Table 2.1). 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 2.1, 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-impoverished condition. In the Bayesian learning model, simulations of the reward-

rich condition resulted in the predicted pattern, while simulations of the reward-impoverished 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 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 paper 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-impoverished 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 towards 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-impoverished condition. In what follows, we empirically test these same predictions.

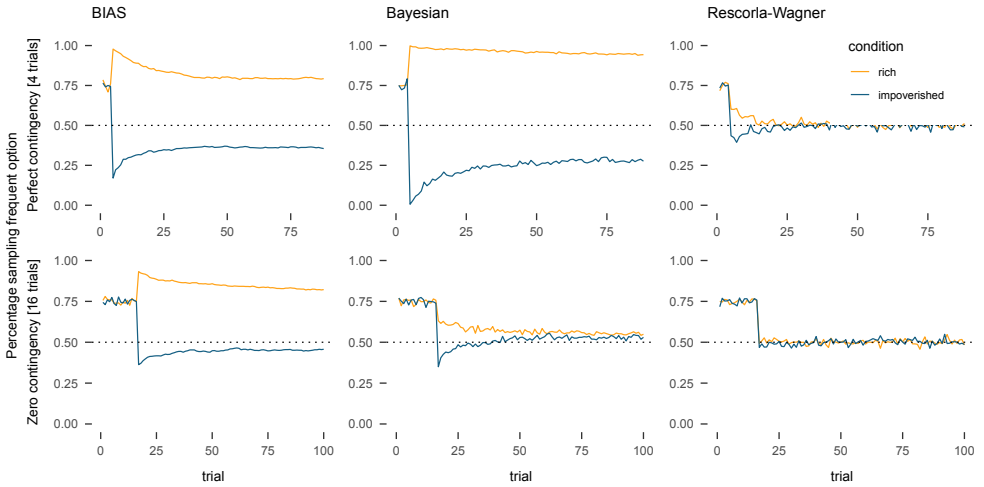



*Figure 2.1.* Probability of choosing the frequent option over the alternative in a reward-rich (green) and a reward-impoverished (orange) condition for the three learning models BIAS, Bayesian learning, and Rescorla-Wagner (left to right). The distribution of initial evidence is identical to that used in Experiment 1 in the first row and identical to that of Experiment 2 in the second row.

## Experiment 1: Perfect Contingency Initial Evidence

Experiment 1 set out to validate the experimental setup and the main assumption underlying this paper. 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-impoverished 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 et al., 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 et al. (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 et al., 2012).

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.

## Methods

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, 2020). One hundred participants ($N_{female}$ = 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 paper was conducted according to the guidelines of the Ethics Review Board of the Faculty of Social & 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, participants encountered initial evidence before then sampling freely. One of the two colors would result in the participant winning ten 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-impoverished 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.

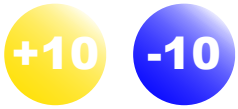
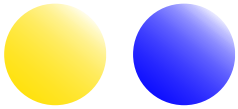
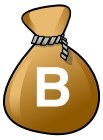
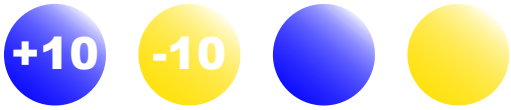
In the induction phase, participants were told that the computer would randomly determine which bag was to be chosen so as to get participants familiar with the task. To this end, they would then only see the one available option. For example, they would see bag A, and, after clicking on this option 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 one second 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 towards 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-impoverished 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-impoverished condition, respectively (Table 2.1).

*Table 2.1*

Distributions of initial evidence

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

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 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-impoverished condition), all measures for this participant were recoded such that a positive value on the relative contingency estimate measure would indicate a bias towards 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 binominal test for the relative contingency estimate that hopefully reduces some ambiguity as to the exact interpretation of the slider[3]. We deem a binominal 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 that the proportion would not be different from .5 in the reward-impoverished condition.



Data preparation and analyses was undertaken using R (R Core Team, 2018) and especially the packages 'dplyr' (Wickham et al., 2019), 'BayesFactor' (Morey & Rouder, 2018), 'lme4' (Bates et al., 2015), and 'lmerTest' (Kuznetsova et al., 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-impoverished 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-impoverished condition, which again warrants one-tailed testing ($BF_{+0}$). While we predict attenuation in the reward-impoverished condition, the mean bias of participants may fluctuate slightly around chance level. For example, we expect relative contingencies to have attenuated towards chance level in the reward-impoverished 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 et al., 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 $H_1$ (Hoijtink et al., 2018). All Bayesian tests use the default

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.

prior of the BayesFactor package, namely a Cauchy distribution of width $r$ = . In appendix 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-impoverished 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-impoverished 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 2.2 and 2.3.

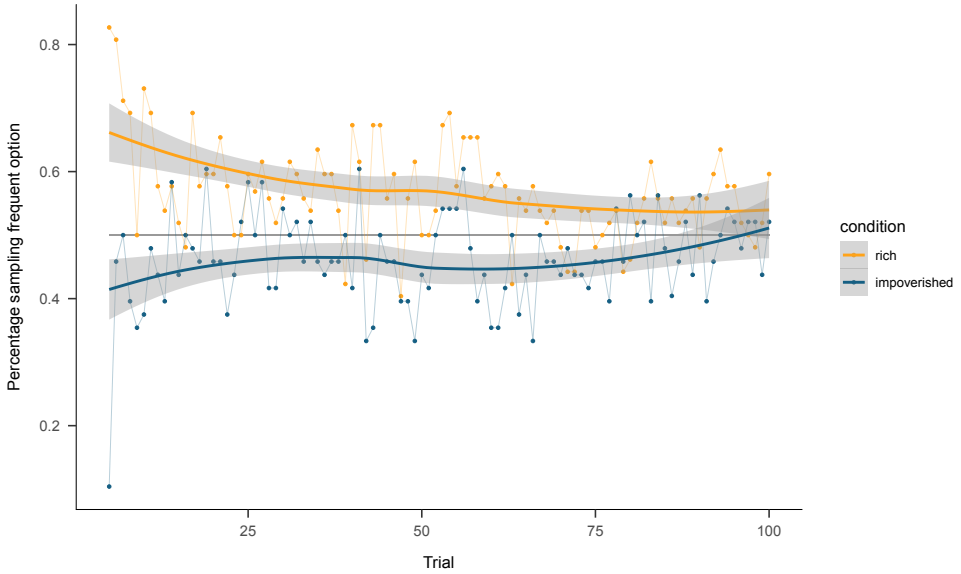



*Figure 2.2.* Percentage of participants sampling the frequent option per trial (Experiment 1).

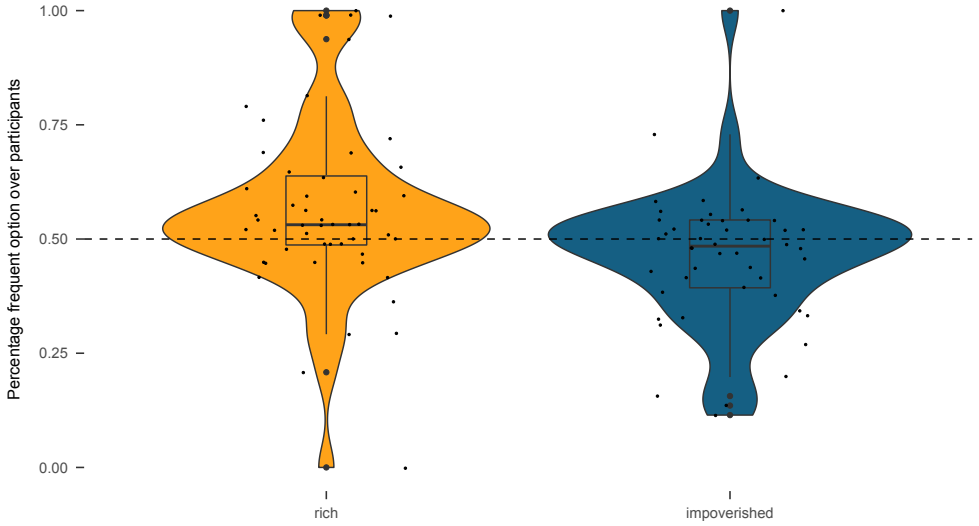



*Figure 2.3.* Proportion of choosing the frequent option per participant (Experiment 1).

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 model[4] to the data in which the reward-impoverished condition is our control and trial number is scaled:

$$\hat{y} = c + \frac{trial}{100} \cdot \beta_{trial} + condition \cdot \beta_{condition} + \frac{trial}{100} \cdot condition \cdot \beta_{trial*condition} \quad (2.3)$$

First, the intercept was not significant which can be seen as testament to how quickly participants in the reward-impoverished 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 towards chance-level in the reward-impoverished 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-impoverished condition as indicated by the non-significant 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 non-significant main effect of condition, $\beta_{condition} = 0.16$, $z = 0.75$, $p = .457$.

We hypothesized that the attenuation in the reward-impoverished 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-impoverished 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-impoverished condition. The remaining trials differ less drastically as can be seen in Figure 2.4. 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.

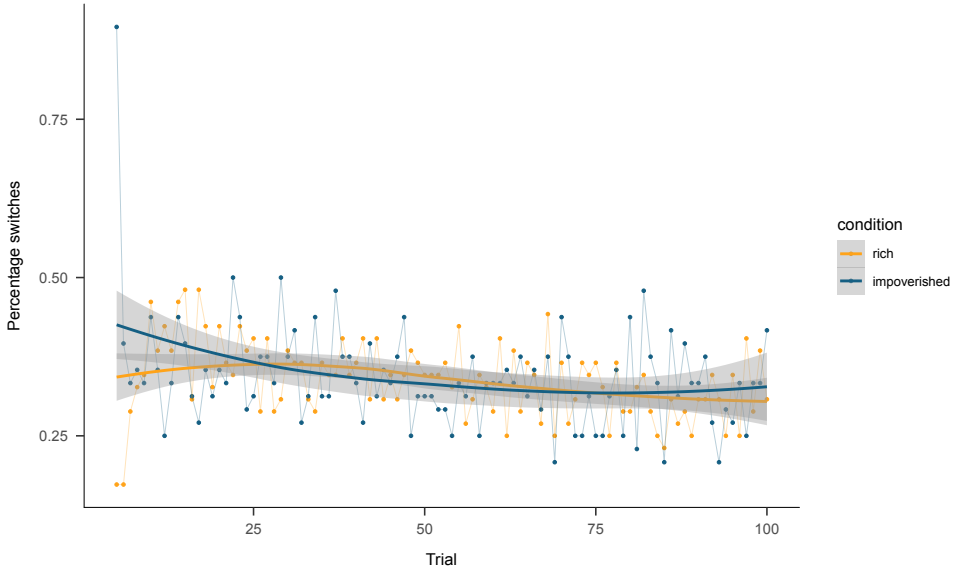



*Figure 2.4.* Percentage of participants switching from previous choice including trend lines per condition (Experiment 1).

***Relative Contingency Estimate***

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 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 $\Delta P$-scores (Allan, 1980). The mean $\Delta P$-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 $\Delta P$-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 judgements 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-subject 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 tradeoff 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 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-impoverished condition.

**Methods**

We decided to double the sample size compared to Experiment 1 so as to rule out any potential problems concerning statistical power. Two hundred participants ($N_{female}$ = 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, 2020).

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 (Table 2.1). The order of the sixteen 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 +/- 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 2.1). 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 Estimate Pre-Measure***

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 towards 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-impoverished 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 Pre-Measure***

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 $\Delta P$-score in the reward-rich ($\Delta P$ = .12, $SD$ = 0.35) and reward-impoverished ($\Delta P$ = .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 Pre-Measure***

Participants estimated their confidence similarly in the reward-rich ($\Delta P$ = .05, $SD$ = 0.23) and the reward-impoverished 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-impoverished condition as can be seen in Figures 2.5 and 2.6. 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-impoverished 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-impoverished 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].

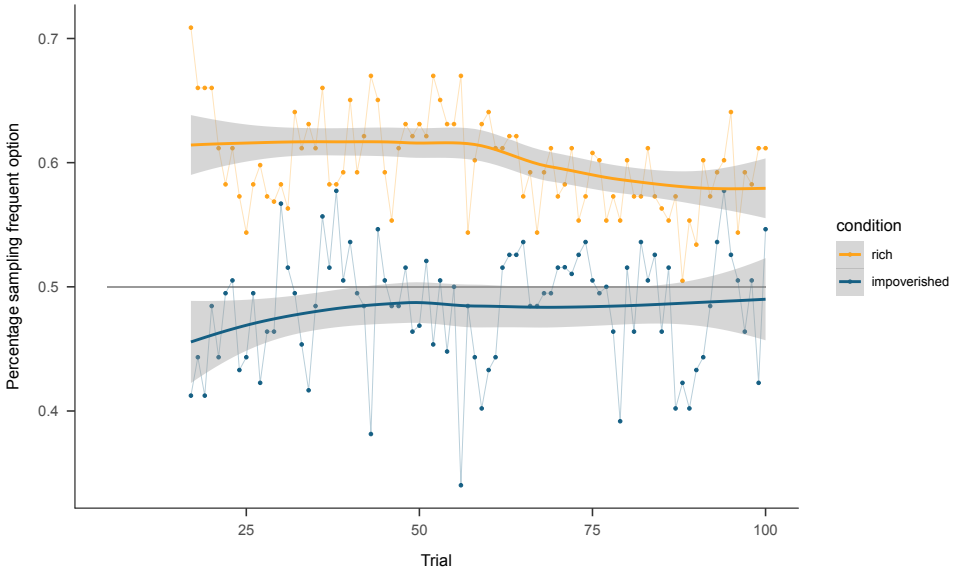



*Figure 2.5.* Percentage of participants sampling the frequent option per trial (Experiment 2a).

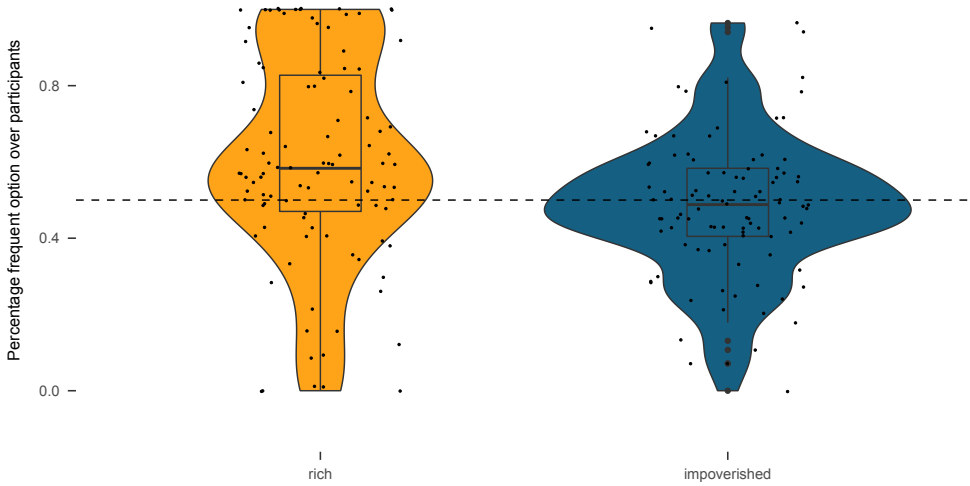



*Figure 2.6.* Proportion of choosing the frequent option per participant (Experiment 2a).

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 non-significant intercept indicates how quickly participants in the reward-impoverished 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-impoverished condition, $\beta_{trial}$ = 0.10, $z$ = 1.04, $p$ = .299. The interaction term again indicates the difference between conditions to decrease over time, $\beta_{trial*condition}$ = -0.45, $z$ = -3.10, $p$ = .002. Once again shifting the model to analyze the last trial, we found attenuation in the reward-impoverished 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-impoverished 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 to 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-

impoverished and 29.1% of participants in the reward-rich condition switched, as can be seen in Figure 2.7.

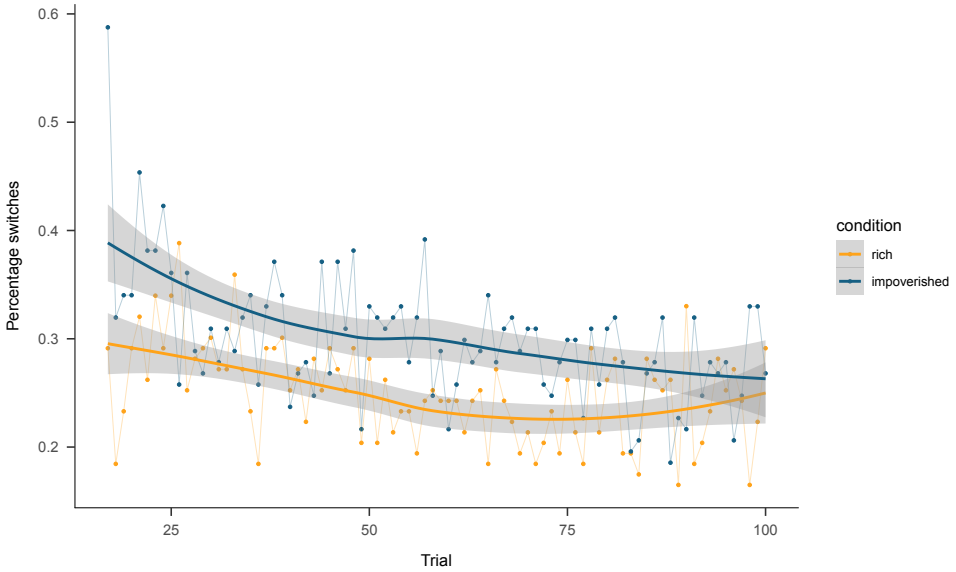



*Figure 2.7.* Percentage of participants switching from previous choice including trend lines per condition (Experiment 2a).

***Relative Contingency Estimate Post-Measure***

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-impoverished condition. Indeed as can also be seen in Figure 2.8, 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-impoverished 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 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 finding support for a difference between the two conditions, $BF_{+0}$ = 8.73, $p$ = .005.



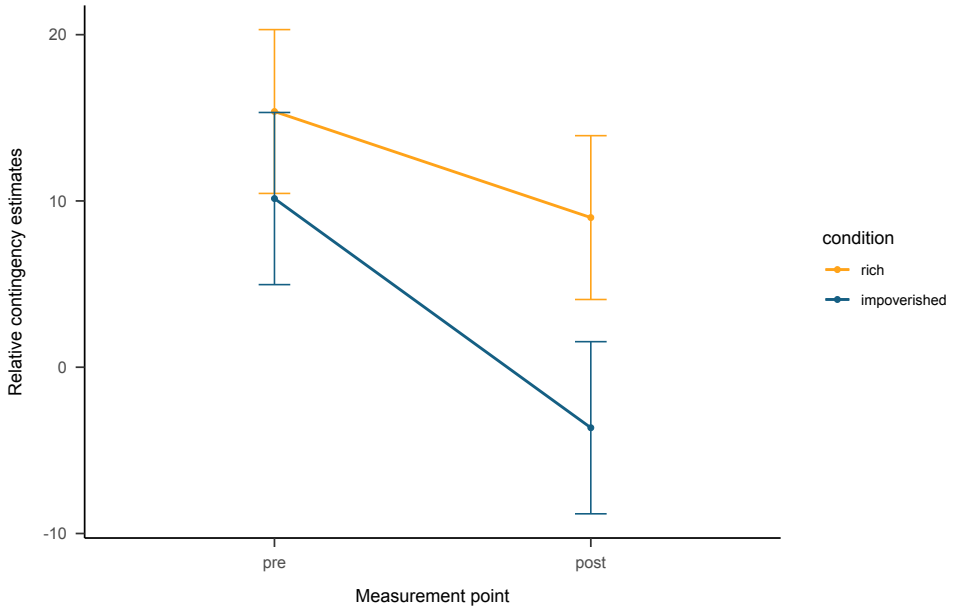



*Figure 2.8.* Relative contingency estimates for the pre and post measurement for the frequent over the infrequent bag (Experiment 2a).

***Conditional Probability Estimates Post-Measure***

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 $\Delta P_{rich}$ = .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 $\Delta P_{impoverished}$ = -.01 (*SD* = 0.25). This is more likely 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 2.9).

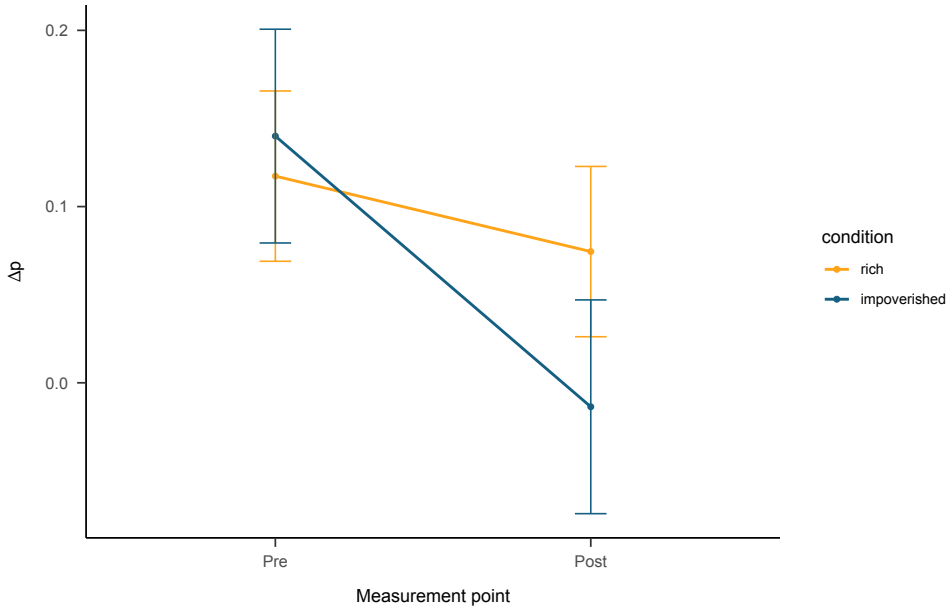



*Figure 2.9. ΔP*-scores from conditional estimates (Experiment 2a).

***Confidence Post-Measure***

Finally, participants again did not estimate their confidence in their judgements 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 pre-measurement across all measures speak to the successful induction of a bias by 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-impoverished 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 afterwards in relative contingency and conditional probability estimates. We found consistent support for the attenuation of any biases in the reward-impoverished 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 (c.f. 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

### Methods

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 ($N_{female}$ = 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 post measure, 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 Pre-Measure*

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 towards 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 frequently shown bag out of the 97 participants in this condition. In the reward-impoverished 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 Estimates Pre-Measure***

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):

$$log_{BR} = log_{10}\left(\frac{ab}{cd}\right) \times log_{10}\left(\frac{ac}{bd}\right) \qquad 2.4$$

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 ($log_{rich}$ = .12, *SD* = 0.16) and reward-impoverished ($log_{impoverished}$ = .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 0 ($BF_{+0}$ > 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 Pre-Measure***

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 ($\Delta P_{rich}$ = .16, *SD* = 0.32) and reward-impoverished ($\Delta P_{impoverished}$ = .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 Pre-Measure***

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.

Participants estimated their confidence similarly in the reward-rich ($\Delta P$ = .03, $SD$ = 0.20) and the reward-impoverished 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-impoverished condition as can be seen in Figures 2.10 and 2.11. 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-impoverished 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-impoverished 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].

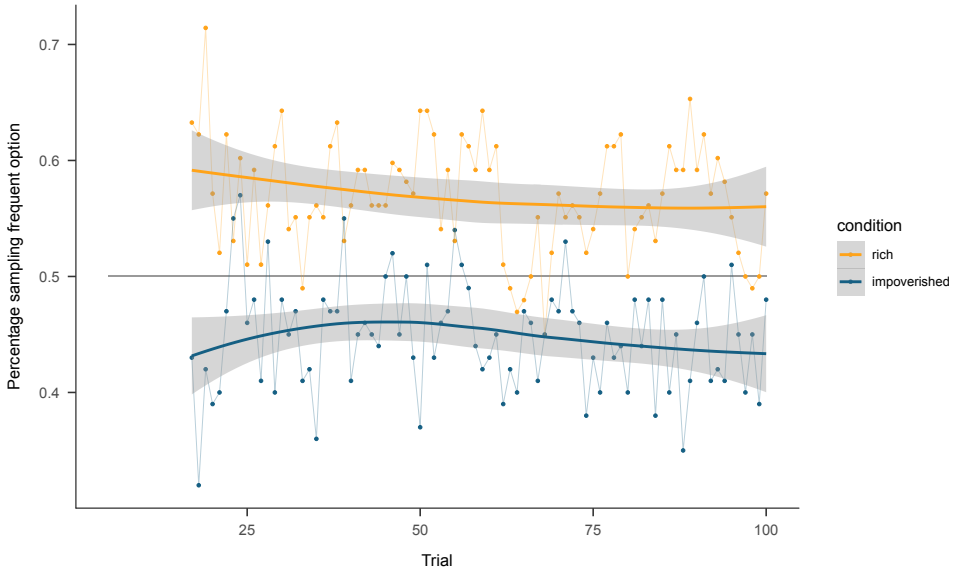



*Figure 2.10*. Percentage of participants sampling the frequent option per trial (Experiment 2b).

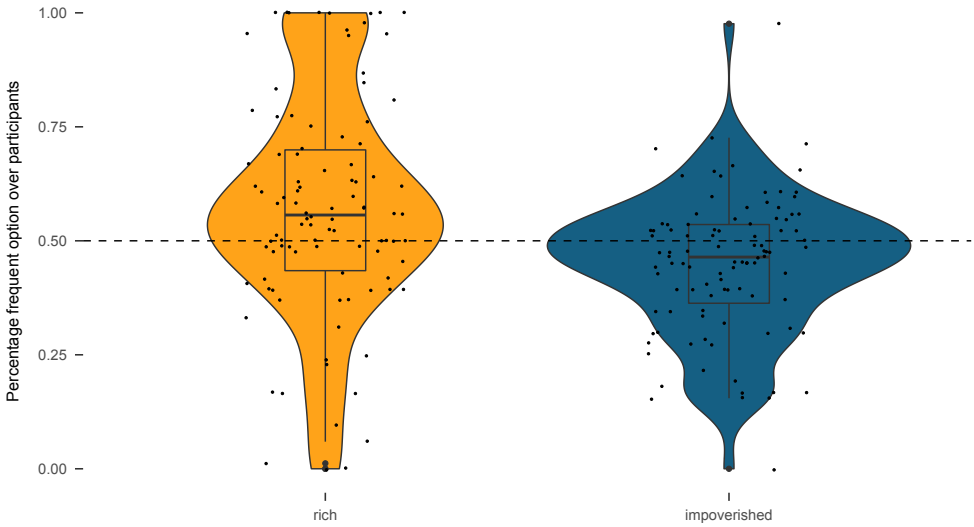



*Figure 2.11*. Proportion of choosing the frequent option per participant (Experiment 2b).

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 non-significant intercept indicates the quick attenuation to chance level in the reward-impoverished condition, $c$ = -0.20, $z$ = -1.55, $p$ = .121. The main effect for condition indicates a successful bias induction in the reward-rich condition, $\beta_{condition} = 0.71$, $z = 3.83$, $p < .001$. The negative estimate for trial, however, indicates not attenuation but a strengthening of the bias across trials in the reward-impoverished condition. The non-significant interaction term, finally, suggests that the difference between conditions did not change over time, $\beta_{trial*condition}$ = -0.10, $z$ = -0.69, $p$ = .488. Finally, shifting the trial number indicated both a lingering bias in the reward-impoverished condition ($c$ = -0.32, $z$ = 2.23, $p$ = .026) as well as the reward-rich condition, $\beta_{condition} = 0.62$, $z = 3.16$, $p = .002$.

We again found the predicted effects on oscillation between the two options. Participants in the reward-impoverished 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-impoverished and 36.7% of participants in the reward-rich condition switching on the first trial, as can be seen in Figure 2.12.

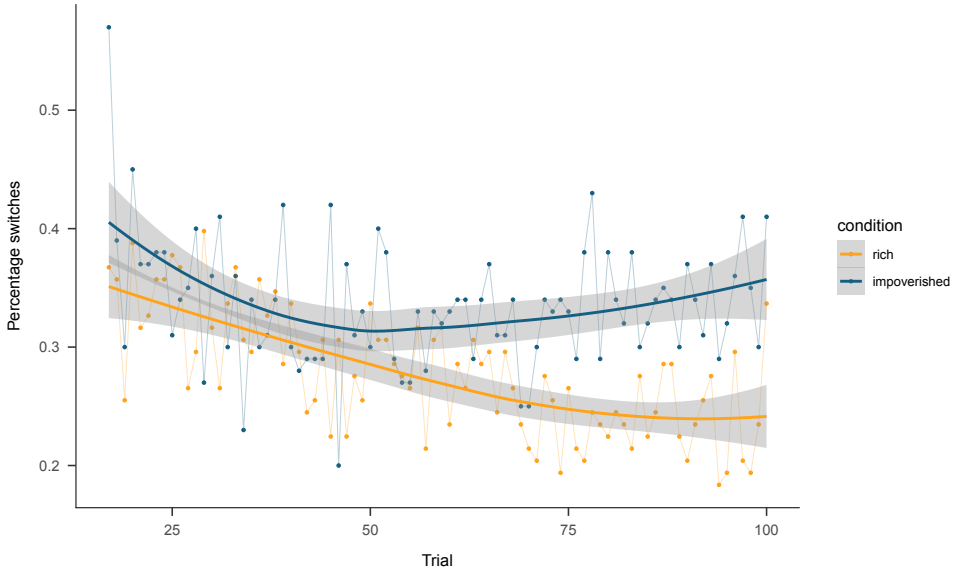



*Figure 2.12.* Percentage of participants switching from previous choice including trend lines per condition (Experiment 2b).

***Relative Contingency Estimate Post-Measure***

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-impoverished condition. Indeed as can also be seen in Figure 2.13, 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-impoverished 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].

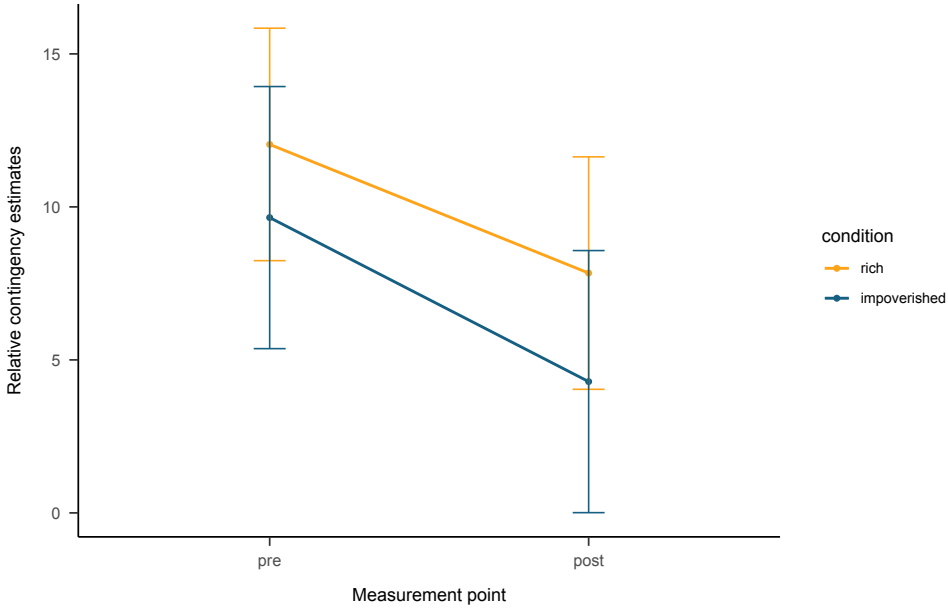



*Figure 2.13.* Relative contingency estimates for the pre and post measurement for the frequent over the infrequent bag (Experiment 2b).

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-impoverished 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-impoverished condition finding no support for a difference between the two conditions, $BF_{+0} = 0.21$, $p = .558$.

***Base Rate Estimates Post-Measure***

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 participants in the reward-rich condition, these estimates resulted in a mean log-score of $log_{rich} = .13$ ($SD = 0.45$), which is about twelve 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 $log_{impoverished} = -.06$ ($SD = 0.30$), 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 2.14. 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.



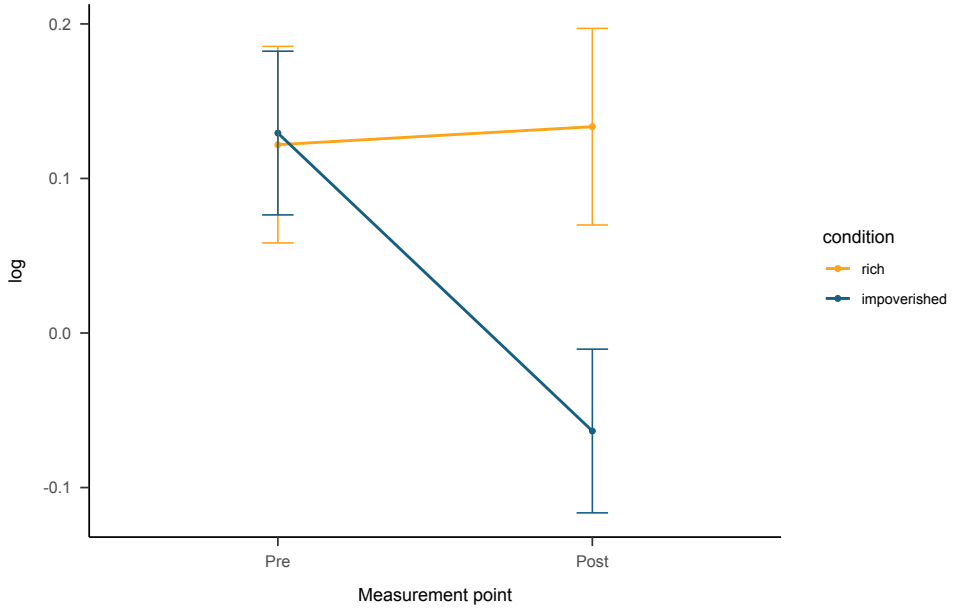



*Figure 2.14. Log*-scores from base rate estimates (Experiment 2b).

### *Conditional Probability Estimates Post-Measure*

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 $\Delta P$-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 $\Delta P$-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 2.15.

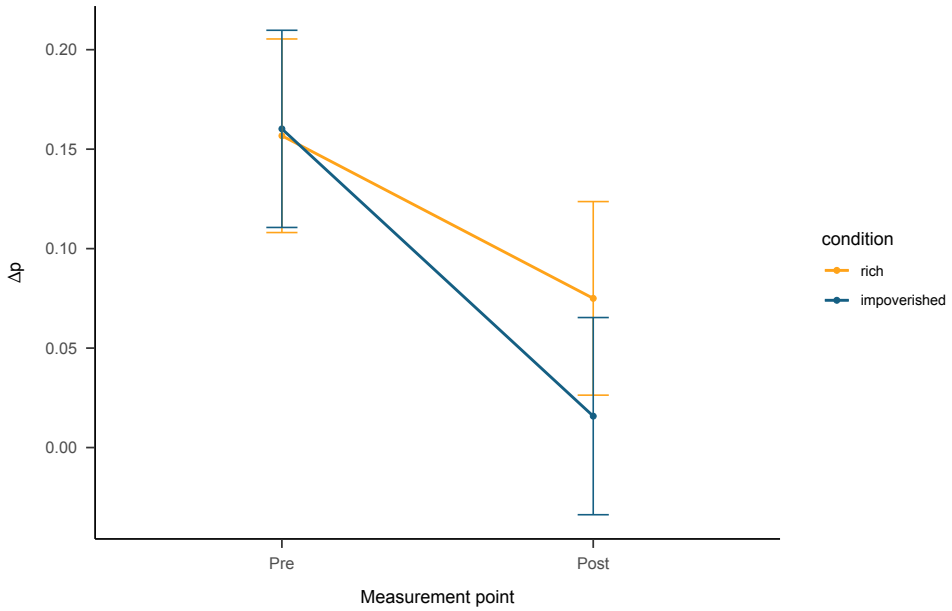



*Figure 2.15. ΔP*-scores from conditional estimates (Experiment 2b).

***Confidence Post-Measure***

Finally, participants again did not estimate their confidence in their judgements to be different in the reward-rich ($\Delta P$ = .02, $SD$ = 0.28) compared to the reward-impoverished ($\Delta P$ = .01, $SD$ = 0.15) condition, $BF_{01}$ = 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 towards 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 to 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 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 to the reward-impoverished 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-impoverished 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-impoverished 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 et al., 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-impoverished 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 et al., 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-impoverished 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-impoverished 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-impoverished 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 towards 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-impoverished 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 confidentially 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 bias[6] 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-impoverished 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.

6 Here, we use distributions with actual contingencies (Experiment 1) and skewed base rates likely to induce pseudocontingencies (Experiment 2a and b). 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

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 towards chance level in the reward-impoverished 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 towards 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 to 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 to, for example, 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 towards 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 towards 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 in environments in which few people are nice towards us would we be likely to readily overcome these initial biases.

Humans and all agentic organisms face the inherent tradeoff 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 tradeoff 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 paper 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 papers we aim to generalize these findings across contexts.

# CHAPTER 3

# Chapter 3:
## When Biases Beget Biases: Erroneous First Impressions of Choice Options Thrive in Settings With Graded Outcomes

**This chapter is based on:**

Harris, C., Fiedler, K., & Custers, R. (2021). When Biases Beget Biases: Erroneous First Impressions of Choice Options Thrive in Settings With Graded Outcomes. *Manuscript submitted for publication.*

## Abstract

While initial inferences are an important determinant of future behavior, they can be inaccurate – which can have downstream consequences. Earlier research has demonstrated that a bias favoring one of two choice options may persist when generally high reward rates seduce people to exploit the seeming advantage of the preferred option and prevent them from exploring the other option's equally high reward rate. However, this differential influence of exploitation and exploration on the persistence of cognitive biases was so far demonstrated only for binary outcomes (wins or losses) and with early judgments committing participants to initial biases. As outcomes in real life are rarely binary and not always subject to self-consistency commitment, the present research extends the paradigm to more meaningful everyday environments with graded outcomes. In three experiments using a modified paradigm with two options leading to graded outcomes, participants were induced to infer a (illusory) contingency favoring one option, before sampling freely between both options. In line with earlier research, the induced initial preferences were either maintained or corrected, depending on participants' sampling strategies. These findings testify to the generality of the phenomenon, which is not peculiar to the restrictive task settings of previous research. They increase our understanding of the conditions under which people maintain unwarranted preferences in continued interaction with the environment despite the availability of counterevidence.

Motivated actions aim at realizing desirable outcomes or goals (Custers & Aarts, 2010; Gollwitzer & Moskowitz, 1996; Kruglanski et al., 2002). While some actions are firmly associated with their outcomes (Aarts & Dijksterhuis, 2000; Bargh, 1990) because a clear-cut causal relation exists (e.g., flicking a light switch turns on the light), people must often learn the dependency of outcomes on their actions as they interact with the world in which they live (e.g., that only visiting a particular deli yields a great pastrami sandwich or that going on a first date is a lot of fun). In these cases, predictions about the outcomes of subsequent actions (visiting that deli or going on a second date) may either be verified or not. But the experience of behavioral outcomes depends on what actions people engage in. Operant learning about action outcomes is not a passive process, but rather an interactive and reiterative one, in which people acquire knowledge about the environment as a result of their goal pursuits (Harris, Aarts, et al., 2022a). Previous research has demonstrated that initial biases are not always attenuated in this reiterative process, but that they can be perpetuated despite plenty of opportunities to put one's beliefs to the test (Harris et al., 2020). Here we build on this research by demonstrating that those findings hold true in more meaningful task settings.



**Exploitation and the Maintenance of Biases**

While the reiterative updating of beliefs about action-outcomes should normally lead to an accurate representation of the world (Behrens et al., 2007; Courville et al., 2006; Yu, 2007), there are also phenomena that suggest that biases might be maintained despite repeated interaction with the environment. One such example is positive testing (Fiedler et al., 1999; Klayman & Ha, 1987; Oaksford & Chater, 1994), which occurs when people search for information that verifies, rather than falsifies a particular hypothesis. Positive testing, however, assumes a dedicated information search. We believe that more often than not, people's actions are driven not by the goal of explicit learning about the environment, but instead by goals such as maximizing rewards (e.g., earning money in an experiment, a good coffee, or even a suitable romantic partner; Denrell, 2005; Thorndike, 1927). Learning, then, often follows in the slipstream of the hedonic goals that drive people's behavior and the evidence they encounter.

This difference is nicely described by the tradeoff between exploration and exploitation (Cohen et al., 2007; Mehlhorn et al., 2015). Exploration is characterized by switching between choice alternatives in order to learn about them. While it allows for learning about these alternatives, it typically does not result in reward-maximization as inferior options are also chosen. Exploitation, on the other hand, is characterized by focusing on a supposedly best option. In this case rewards are maximized, while learning about alternatives is reduced or absent. Exploration and exploitation form

a continuum and while decision-makers rarely engage in its most extreme forms (cf. Fantino & Esfandiari, 2002; Gaissmaier & Schooler, 2008; Vulkan, 2000) this framework nicely emphasizes how people learn under uncertainty (Gershman & Daw, 2017; Hogarth et al., 2015) and from (previous) experiences (Denrell, 2005; Denrell & Le Mens, 2011; Denrell & March, 2001; Gilboa & Schmeidler, 1995; Hertwig & Erev, 2009). Iterative choices must solve a tradeoff between adaptive learning and hedonic goal-fulfillment.

In a recent series of experiments, Harris et al. (2020) demonstrated that the balance of exploration and exploitation is strongly influenced by the outcomes encountered during sampling. Whereas environments with an overall high winning rate facilitate exploitation of profitable strategies, environments with an overall low winning rate facilitate exploration and strategy shifts. This resulted in people sticking to options that seemed attractive initially in the reward-rich, but not in the reward-impoverished environment. Specifically, participants played a two-armed bandit task in which they had the opportunity to choose between two options that yielded either a 10-point win or a 10-point loss with points being converted to a financial reward at the end of the experiment. In an initial learning stage, participants were exposed to a double-skewed stimulus distribution that previous research had shown to regularly induce a pseudocontingency illusion (Fiedler et al., 2009), creating a bias in favor of one over the other option. That is, when the same high rate of 75% (or low rate of 25%) wins holds for both options, but one option is chosen more frequently than the other, the frequent winning outcome appears to correlate with the frequent option.[7]

After the induction of such an illusory preference for one option in the initial stage, in a second stage participants could then freely sample observations from both options for the remainder of one hundred trials to earn as many points as possible. Results showed that initial biases persisted, but only when a high overall winning rate encouraged exploitation. In such a reward-rich environment, participants continued to sample more from the frequent than from the infrequent option, thus preventing them from noting the same winning rate of the allegedly inferior option. In contrast, a low winning rate in a reward-impoverished setting motivated participants to engage in exploration, switching from the frequent to the infrequent option, which allowed them to experience the actual zero contingency.

We believe that these demonstrations of how exploitation can serve to maintain initial biases whereas exploration offers a chance to learn about alternative outcomes and hence to correct for unwarranted biases is of fundamental theoretical and practical importance. It helps us to understand, for instance, why people develop

7 As the infrequent outcome also appears to come along with the infrequent option, the pseudocontingency illusion cannot be reduced to a bias toward the most frequent stimulus combination.

idiosyncratic beliefs about certain behaviors (e.g., belief in superstitious rituals, trust in alternative medical treatments, preferences for certain foods) and why failure experience can be important for adaptive behavior regulation.

However, the paradigm and the evidence reported by Harris et al. (2020) are incomplete and call for further elaboration in several ways. First, the findings may be restricted to distinct features of the original task setting. For instance, the two-armed bandit task setting resembles an abstract statistical inference task that was restricted to dichotomous outcomes (either winning +10 or losing -10), whereas real outcomes are often graded. Another distinct feature of the original task setting is that the initial bias-induction stage was clearly separated from the sampling stage by procedure and explicitly fixated by a final contingency judgment, to which participants may have been then committed. Second, pseudocontingencies are by no means the only cause of an initial bias that can trigger different strategies of information search. Different biases can affect information search in different ways. And third, the effects of exploitation and exploration can be examined at different aggregation levels. Reward-rich versus reward-impoverished environments may be analyzed at the level of a general tendency to exploit or to explore across the entire sampling period. But this tendency can also be understood as the result of distinct hedonic influences at the level of specific trials.

***Peculiarity to Task Settings.***

In an attempt to improve the external validity and the social relevance of the paradigm, the present research sought to investigate the influence of exploitation and exploration on the persistence of biases in a modified task setting. In the modified paradigm, outcomes are not restricted to dichotomous format, wins versus losses, but appear in graded format. A simple and concrete cover story (i.e., comparing fruit sales in different places) renders the contingency learning task meaningful and distinct from an abstract statistical inference task.

Moreover, the impact of the initial preference-formation treatment may depend on whether an explicit contingency judgment task at the end of the initial stage actually binds the more frequent outcome to the more frequent option (as in an attention-shift paradigm; Kruschke, 2003) and commits the participants' sampling strategies to their explicitly stated preferences. A logical premise of the basic expectation that an initial bias is maintained in a reward-rich (but not in a reward-impoverished) environment is of course logically dependent on the successful induction of an initial bias, which depends in turn on a specific frequency pattern of both options' wins and losses experienced in a distinct time segment. Thus, the inclusion versus exclusion of an initial contingency judgment not only offers a check

on the success of the bias-induction manipulation. It may in fact be a necessary requirement to render the manipulation effective.

***Sources of Bias***

The basic idea that exploitation and exploration can stabilize or undo an initial bias, respectively, not only applies to pseudocontingencies but to other sources of bias as well. Recall that a pseudocontingency illusion arises when the frequency distributions of outcomes and options are skewed in the same direction. That is, a frequent outcome is linked to the frequent option and, conversely, an infrequent outcome is linked to the infrequent option. According to this algorithm, a reward-rich setting (with, say, 9 wins / 3 losses for the frequent and 3 wins / 1 loss for the infrequent option) should induce an illusory preference for the frequent option, whereas a reward-impoverished setting (3 wins / 9 losses and 1 win / 3 losses) should induce an illusory preference against the frequent option.

However, rather than showing a pseudocontingency illusion, based on the alignment of two skewed frequency distributions, participants may instead fall prey to another well-established frequency illusion, the so-called ratio bias (Denes-Raj et al., 1995; Reyna & Brainerd, 2008). When the winning to sample-size ratio is held constant, people would typically prefer the larger lottery or sample, which offers a larger absolute number of winning opportunities. Accordingly, a ratio bias predicts a preference for 9/(9+3) over 3/(3+1), but also a preference for 3/(9+3) over 1/(3+1). In either case, the larger sample offers more opportunities to win. In other words, in a reward-rich setting, both ratio bias and pseudocontingency induce a preference for the frequent option, but in a reward-impoverished setting, the ratio bias works against the pseudocontingency bias. Whereas the latter induces a preference for the infrequent option, the ratio bias always produces a preference for the more frequent option.

Thus, different participants who, for any reason, fall prey to different cognitive illusions, can be expected to exhibit different rates of exploitation and exploration. Specifically, our refined paradigm predicts that, regardless of what bias is at work in a reward-rich environment, a successfully induced preference for the frequent option should facilitate exploitation and therefore serve to maintain the initial bias. In contrast, in a reward-impoverished setting, the kind of bias that exploitation and exploration strategies can be expected to conserve or correct, respectively, depends on what bias dominates the initial preference induction process. In any case, a reward-rich setting should facilitate sampling from, and hence maintaining a preference for, the initially preferred option.

***Levels of Exploration and Exploitation***

We not only modified and refined the design and procedures of the Harris et al. (2020) paradigm but also the methods of statistical analysis. So far, we only considered the extent to which the hedonic structure of a (reward-rich vs. – impoverished) task setting induced a general tendency to exploit or to explore, leading in turn to the conservation or correction of an initial bias, respectively. However, such a general tendency to exploit or explore, across all trials, is by no means the only way in which hedonic influences affect behavior. In the modified and refined paradigm, we not only analyze the overall trend to exploit or to explore but also the flexibility and impulsivity, with which participants change their sampling focus. Using the "win-stay-lose-shift" (Nowak & Sigmund, 1993) rate as a measure of the extent to which the option sampled on trial $t + 1$ is contingent on the outcome on trial $t$, we analyze the course of the sampling process in reward-rich and reward-impoverished environments on a trial-by-trial level. Would the enhanced tendency to maintain an initial bias in the reward-rich (compared to the reward-impoverished) condition reflect a monotonic sampling preference for the frequent option, which remains constant over trials? Or would the sampling bias toward the frequent option reflect the participants' sensitivity and reactivity to hedonic outcome value? Note that a tendency to sample predominantly from the frequent option need not reflect an invariant sampling preference. It may as well reflect an impatient, frequently shifting sampling style that follows the hedonic rule of win-stay-lose-shift. Thus, a reward-rich schedule may make participants less tolerant and more loss-averse, so that they frequently shift to the other option after a loss trial. Yet, given the higher win than loss rate in a reward-rich setting, such an impulsive strategy provides more reasons to stay with the modal option than to shift. In any case, a reward-rich environment may not only induce conservative and bias-affirming strategies but also impulsive and highly reactive styles.

Here, we investigate the influence of reward-rich and reward-impoverished environments on the maintenance or attenuation of initial biases in a task with more naturalistic parameters. The three experiments reported in the remainder of this article will cover possible variants of initial preferences acquired in the first stage (e.g., with or without inviting self-consistency) and different sampling strategies (exploitation vs. exploration) in the second stage. Experiment 1 uses a distribution of initial evidence that was successfully used in previous studies (Experiment 2; Harris et al., 2020) meant to induce pseudocontingencies. The outcomes are, however, graded and we do not include a probe for contingency estimates after the induction phase. In Experiment 2 we return to these explicit contingency estimations before the sampling phase. Experiment 3 then uses a distribution that contains

an actual, perfect, contingency. We believe the results provide new insights on preference formation in decision-making tasks with repeated choices. Specifically, they demonstrate that depending on the environment and depending on the style of processing, unwarranted preferences for choice options based on first impressions can be upheld despite continuous interaction with the choice alternatives. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the experiments (Simmons et al., 2012)[8].

**Experiment 1**

Let us first introduce the experimental setup that modifies the general setup used by Harris et al. (2020) in several distinct ways. In the present paradigm, participants were instructed to sell strawberries from either of two locations. Instead of binary outcomes, outcomes were now gradated with variance in how much one had earned on any given trial. Importantly, we told participants that there was an (artificial) threshold framed as "breakeven point". Depending on this breakeven point, participants should interpret a trial either as a success or as a failure. To increase external validity, we did not ask participants to actively focus on the underlying distribution of evidence encountered by explicitly asking them for contingency estimates following the induction phase. In line with previous research, we expected participants to form a bias in favor of one location, based on a pseudocontingency illusion triggered by the initial evidence. They should then maintain this initial bias in the reward-rich condition but attenuate the bias in the reward-impoverished condition. An a priori power analysis for a difference from constant *t*-test using G*Power (Faul et al., 2007) based on effect sizes in Experiment 2 of Harris et al. (2020) suggested a minimum sample size of at least 50 participants per condition to be necessary. These calculations were based on a 5% alpha-level, 80% statistical power, and effect sizes between $d = .35$ and $d = .43$ as reported by Harris et al. (2020) for maintained biases in the reward-rich condition. A sensitivity analysis, also using G*Power, ($\alpha = 0.05$, $1\text{-}\beta = .80$, $N = 201$) indicated that we could detect effect sizes as small as $d = 0.18$.

**Methods**

Participants were recruited via the online crowdsourcing platform Prolific Academic. The experiment was run in English on SosciSurvey (Leiner, 2020). Two hundred and one participants ($N_{female} = 145$) with an average age of 33 years ($SD = 8.93$) participated for a financial reward of at least £0.85 plus additional performance-dependent earnings (max £2.15, mean £1.50). The research line reported in this

8 The order in which we ran the experiments is Experiment 3, Experiment 1, Experiment 2. The order here represents our best explanation regarding the mechanisms at work. We are careful to avoid the term 'replication' where it is not befitting and always report our originally hypothesized results first. Additional post hoc analyses are clearly labeled as such.

article was conducted according to the guidelines of the Ethics Review Board of the Faculty of Social and Behavioral Sciences at Utrecht University.

***Design***

The experiment used a two-armed bandit task with one hundred trials. Participants were asked to imagine wanting to earn some pocket money by selling strawberries. They could choose either of two locations (Figure 3.1) leading to positive or negative outcomes, symbolically represented by the number of strawberry baskets they sold (Figure 3.2). Participants were assigned to either a reward-rich or a reward-impoverished condition and depending on their assignment would end up selling on average either many or few strawberry baskets. We counterbalanced which location was shown more often in the initial evidence phase.



*Figure 3.1*. The two locations participants could choose: a street corner and a park.

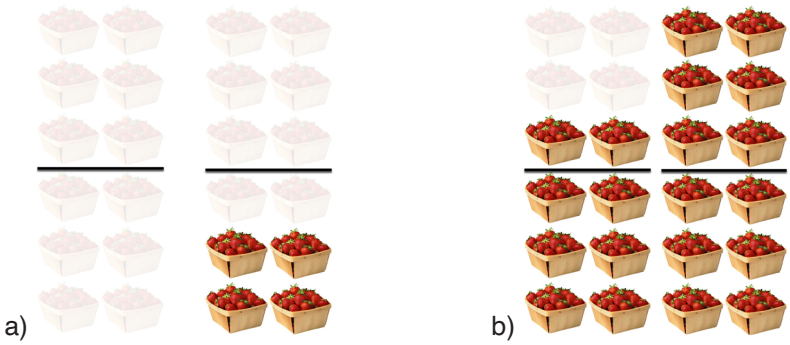



*Figure 3.2*. Examples of the outcomes participants could encounter on any given trial. Depicted are the lowest and highest number of baskets for loss trials (a) and win trials (b) as categorized by the “breakeven”-line.

***Procedure***

First, participants were introduced to the task. They were instructed to imagine selling strawberries over a summer. They were asked to consider two locations from which they could sell strawberries, a street corner and a park. The outcomes were symbolically depicted as more or less strawberry baskets supposed to indicate successful versus unsuccessful days. They were reminded that the task was incentivized, and that they should try to sell as many strawberries as possible in order to maximize their earnings for participation.

Then, in the induction phase, participants were told that for the first few trials the computer would randomly choose which of the two locations participants could sell strawberries from in order to get them familiarized with the task. They would then only see (and select) the one available location that was supposedly chosen by the computer. After selecting this location, the outcome of the chosen location would be presented in the form of text ("You sold from location 'street' ['park'] and sold many [few] strawberries") as well as images. These images were the picture representing the chosen location and a number of strawberry baskets symbolically depicting how many strawberries had been sold. This image consisted of 12 baskets and a black line dividing the baskets in the upper six and lower six baskets. Participants were told that this line represented the breakeven point and that everything above the line should be considered a win, but everything below the line should be considered a loss. After a delay of one second, the next selection was presented. During the entire experiment, the current trial number was presented on the screen so that participants knew the remaining number of trials.

The induction phase lasted for sixteen trials (Table 3.1). Participants were forced to select one location more frequently than the alternative and would either win (reward-rich condition) or lose (reward-impoverished condition) more often. The order of these sixteen trials was random.

Table 3.1
*Distributions of initial evidence*

| | Experiment 1, 2 | | | Experiment 3 | | |
|---|---|---|---|---|---|---|
| Reward-rich | Wins | Losses | | Wins | Losses | |
| Frequently shown location | 9 | 3 | 75% | 3 | 0 | 100% |
| Infrequently shown location | 3 | 1 | 75% | 0 | 1 | 0% |
| | | | $\Delta P = 0$ [no contingency] | | | $\Delta P = 1$ [perfect contingency] |
| Reward-impoverished | Wins | Losses | | Wins | Losses | |
| Frequently shown location | 3 | 9 | 25% | 0 | 3 | 0% |
| Infrequently shown location | 1 | 3 | 25% | 1 | 0 | 100% |
| | | | $\Delta P = 0$ [no contingency] | | | $\Delta P = 1$ [perfect contingency] |

*Note*: Distributions used for the initial induction phase. All percentages depict the ratio of wins out of all trials for the respective location. $\Delta P$ is the difference score between the conditional probabilities and describes the contingency between location and outcome (Allan, 1980)

Then, in the free choice phase, participants could choose freely between both locations for the remaining 84 trials. Across all trials both locations were equally likely to result in wins, namely 75/100 trials in the reward-rich and 25/100 trials in the reward-impoverished condition. In the final phase, participants gave estimates regarding their preferences for either location. Specifically, we asked them at which of the two locations they were more successful in selling strawberries, a relative contingency measure. We also asked them to give conditional probability estimates on how likely it was to earn a lot of strawberry baskets by choosing either the street corner or the park location. We asked participants how confident they were in making a reasonable estimate. Finally, we asked separately for both locations how much they felt in control of the outcome when choosing this location and how well they felt they could predict the outcome when choosing this location. These last two measures were not previously recorded in Harris et al. (2020). To reduce the length of the results sections, we only give a brief overview of the relative contingency estimates and conditional probability estimates in text; we provide results for control and predictability for all experiments in appendix B.

All questions were answered by moving sliders. For the relative contingency estimate, participants could move a slider with the two locations as the extremes. All other measures consisted of a slider for the street corner and for the park location. The confidence measure was anchored with “not confident at all” and “very confident”. The conditional probability estimates, control estimates, and predictability estimates were anchored at 0% and 100%.

***Data Preparation***

We recoded the data so that positive values represent a preference for the more frequent location in the induction phase. If a participant won nine times (in the reward-rich; lost in the reward-impoverished condition) after having to choose the street corner in the induction phase, a positive value (e.g., on the relative preference scale) meant they still preferred this location over the (less frequent) alternative. To analyze choice behavior, we coded every choice of the frequent location as +1 and every choice of the alternative as 0, effectively creating a choice index of participants’ overall preference. We then divided this index by the total number of free trials and report here in percentages. The relative contingency measure ranged from [-50; 50]. From the conditional (and confidence) estimates we calculated $\Delta P$-scores, the difference between the conditional probabilities (Allan, 1980; Jenkins & Ward, 1965), with a range of [-1; 1] analogous to the well-known correlation measure.

All data preparation and analyses were done using R (R Core Team, 2018) and all Bayesian analyses were undertaken using the package “BayesFactor” (Morey & Rouder, 2018). Given our directional hypothesis that biases would persist in the

reward-rich condition, we perform one-sided tests ($BF_{+0}$). We did not expect biases to persist in the reward-impoverished condition and perform two-sided tests ($BF_{01}$). We also expected the reward-rich condition to differ from the reward-impoverished condition and we performed one-sided tests ($BF_{+0}$) to test these differences in the absolute deviation from chance level (Nieuwenhuis et al., 2011). That is, we compared whether there was a difference between conditions independent of the direction of the bias and refer to these tests as the strength of the effect. Finally, we report overall effects in which we compared all participants across conditions to chance-level. If not indicated otherwise all tests are *t*-tests and their Bayesian equivalent.



Bayes factors indicate the relative support for a hypothesis (e.g., $H_1$) over a competing hypothesis (e.g., H0; Hoijtink et al., 2018). All Bayesian tests use the default settings of the BayesFactor package, including a Cauchy distribution of width $r =$ . In appendix B, 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. In all graphs the measure of dispersion are confidence intervals, which we also report for the effect sizes, denoted with a subscript d ($CI_d$). All analyses and data can be found on an Open Science Framework repository (https://osf.io/vjz7b/?view_only=01030f9ffd124aa08d878fa21ffd7776).

**Results**

***Sampling***

Figure 3.3 summarizes participants' choices across trials. In a first analysis we examined participants' sampling behavior using a logistic mixed-effects model with trial number and condition as predictors and participants as random effects[9]. Due to the binary outcomes, we fitted the following logistic model to the data in which the reward-impoverished condition was our control (coded as 0) and trial number was shifted, so that the first free choice trial represents the model's intercept, as well as scaled:

$$\hat{y} = c + condition \times \beta_1 + \frac{trial}{84} \times \beta_2 + \left(\frac{trial}{84}\right)^2 \times \beta_3 + \frac{trial}{84} \times condition \times \beta_4 + \left(\frac{trial}{84}\right)^2 \times condition \times \beta_5 \quad (3.1)$$

If the first 16 observations exerted the same influence on the subsequent free choice stage as in three experiments reported by Harris et al. (2020), an enhanced exploration tendency in the reward-impoverished condition should be evident in a reduced tendency to continue sampling from the frequent location, compared to the

9 The graph suggests a non-linear trend and indeed the quadratic model outperformed a linear model, $\chi^2(2) = 7.34$, $p = .025$.

reward-rich condition. However, a non-significant intercept indicated no initial bias in the reward-impoverished condition, $c = 0.07$, $z = 0.65$, $p = .517$. Moreover, although the line in Figure 3.3 for the reward-rich condition proceeds persistently above that of the reward-impoverished condition, the overall difference between the reward-impoverished and the reward-rich condition was not quite statistically significant, $\beta_2 = 0.27$, $z = 1.68$, $p = .093$. Neither trial number nor the Trial number × Condition interaction were significant (see Table 3.2 for the full choice model).

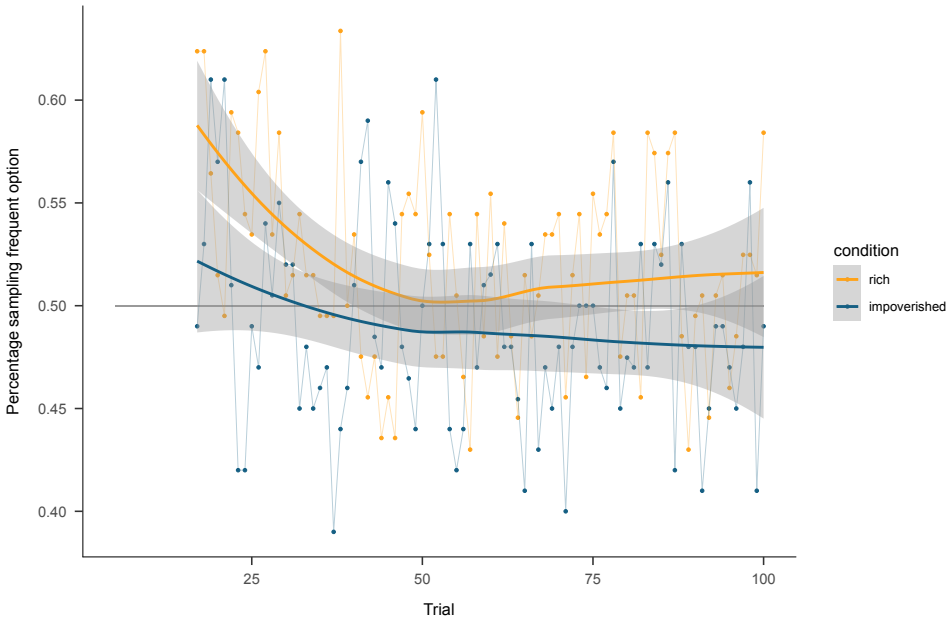



*Figure 3.3*. Percentage of participants sampling the frequent location per trial (Experiment 1). Line graphs display a local polynomial regression fit with 95% CI.

Table 3.2

*Regression table for choice model in Experiment 1*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| Intercept | 0.07 | 0.65 | .517 |
| Condition | 0.27 | 1.68 | .093 |
| Trial | -0.50 | -1.32 | .189 |
| Trial$^2$ | 0.39 | 0.44 | .380 |
| Trial × Condition | -0.74 | -1.37 | .171 |
| Trial$^2$ × Condition | 0.80 | 1.25 | .213 |

*Note.* Quadratic logistic mixed-effects model, trial number and condition as fixed effects, participants as random effects.

The failure to reproduce a constant influence of conditions on a tendency to exploit versus to explore may be due to the omission of a committal contingency estimate after the initial 16 trials, as already noted. It could hardly reflect a ratio bias working against the expected pseudocontingency biases because both biases would clearly predict a preference for the frequent condition in the rich-reward condition. It is of course also possible, but unlikely, that the same treatment that had reliably induced the same sampling bias in prior research turned out to be non-replicable. We were thus inclined to believe in the first option, which can be tested by including in the next experiment a contingency pre-estimation task after the initial phase.



However, although Experiment 1 did not induce a constant sampling difference between reward conditions, it did provide rather strong evidence for another manifestation of the influence of reward-rich versus reward-impoverished settings on exploitation and exploration, respectively. This was evident in an analysis of the tendency to exhibit a win-stay-lose-shift strategy, that is, to repeat choosing the same location after a win ("stay" coded as 0) but to shift to the other location after a loss ("shift" coded as 1). Our analytical model was a logistic mixed-effects linear model with participants as random effects and in which the condition, the binary outcomes and the extremity of the feedback were inserted as fixed effects. The binary outcomes indicated whether the previous trial was a win (coded as 1) or a loss (coded as -1), and extremity indicated how extreme the feedback was (1 = 4 or 8 baskets, 2 = 3 or 9 baskets, up until 5 = 0 or 12 baskets). We included reward-rich (coded as 1) versus reward-impoverished condition (coded as 0) as a potential moderator. Importantly, a model that included extremity of the feedback (coded from 1-5) outperformed models that did not include extremity as a fixed effect, indicating the importance of gradedness on the choices, $\chi^2(4) = 198.22$, $p < .001$. In other words, the gradedness of the outcomes explained a significant part of variance in predicting participants' choices. We report the full mixed-effects models in Table 3.3. Additionally, Figure 3.4 visually presents the means across participants.

In a model that only considered the main effects of condition, the binary outcome, and the extremity of the feedback, we found two significant results, for binary outcomes and extremity of feedback. Shifts (from one location to the other) were more likely after losses than after wins, $\beta = -0.48$, $z = -21.80$, $p < .001$, reflecting an overwhelming win-stay-lose-shift effect. More extreme outcome feedback reduced the tendency to shift, $\beta = -0.04$, $z = -2.71$, $p = .007$.

When interaction terms were included in the regression analysis, a more detailed picture emerged. An interaction between binary outcomes and extremity was due to more extreme wins increasing the probability of staying and more extreme losses increasing the probability of shifting, $\beta = -0.07$, $z = -3.27$, $p = .001$. Apparently,

then, the win-stay-lose-shift principle manifests as a win-more-stay-more-lose-more-shift-more effect.

Most importantly, both the binary outcomes and feedback extremity came to interact with conditions. The tendency to shift more after losses than after wins was weaker in the reward-rich condition and accentuated in the reward-impoverished condition, $\beta = -0.28$, $z = -2.84$, $p = .005$. Decision-affect theory offers a plausible account for this interaction (Mellers et al., 1997). Because infrequent losses (25%) are particularly disappointing, the tendency to shift was stronger in the reward-rich condition than in the reward-impoverished condition, in which as many as 75% losses may have led to habituation. Alternatively, the lesser divergence of the two curves for the reward-impoverished condition (see left part of Figure 3.4) may reflect a ceiling effect; maybe a rate of 75% loses is simply too high to always solicit a choice shift. In any case, although the manipulation of reward conditions failed to induce constant sampling preference, it exerted a distinct influence on the participants' reactivity to the outcome of individual trials.

A two-way interaction reflects that the accentuation of binary outcomes (win vs. loss) in the reward-rich (vs. reward-impoverished) condition increases with feedback extremity, $\beta = -0.23$, $z = -7.56$, $p < .001$, as evident in the divergence of the two curves, particularly on the right side of Figure 3.4. Note, however, that as 75% of all trials were wins in the reward-rich but losses in the reward-impoverished condition, the line graphs referring to these two conditions represent many more trials than the other two line graphs. Finally, all three two-way interactions were qualified by a three-way interaction. Shifting was more likely following losses than wins. This difference became more pronounced with more extreme outcomes and it was more pronounced in the reward-rich compared to the reward-impoverished condition, $\beta = -0.06$, $z = -2.02$, $p = .044$.

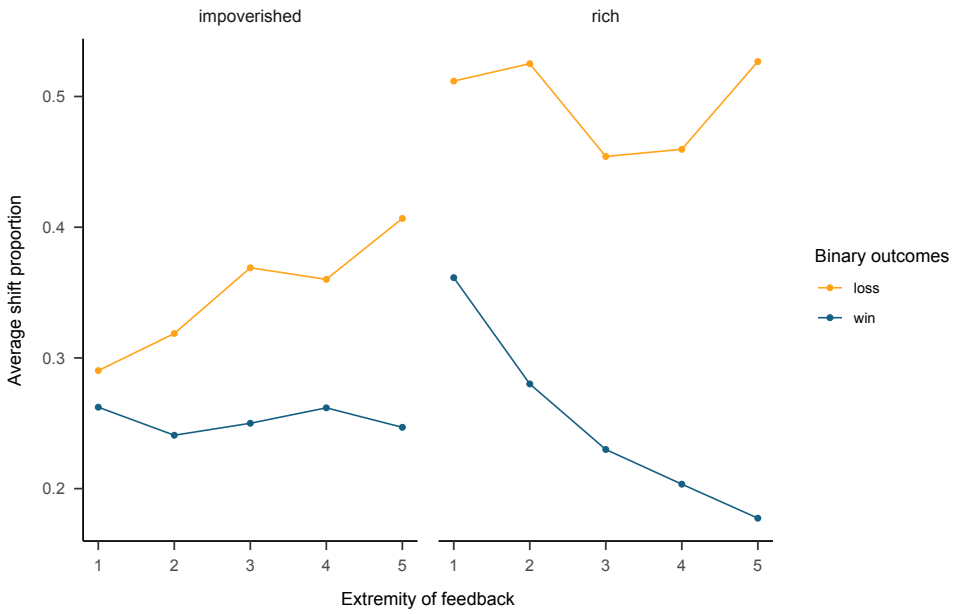



*Figure 3.4.* Average proportion of trials on which participants shift from previous option given the condition (reward-impoverished vs. reward-rich) and extremity of feedback (1 - 5) in Experiment 1.

Table 3.3

*Regression table for shift model in Experiment 1:*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| **Main effects model** | | | |
| (Intercept) | -0.99 | -8.53 | < .001 |
| Condition | 0.28 | 1.76 | .079 |
| Binary outcome | -0.48 | -21.80 | < .001 |
| Extremity | -0.04 | -2.71 | .007 |
| **Interaction effects model** | | | |
| (Intercept) | -1.17 | -8.94 | < .001 |
| Condition | 0.88 | 4.75 | < .001 |
| Bin. Out | -0.04 | -0.50 | .619 |
| Extremity | 0.07 | 3.21 | .001 |
| Condition × Bin. Out | -0.28 | -2.84 | .005 |
| Condition × Extremity | -0.23 | -7.56 | < .001 |
| Bin. Out × Extremity | -0.07 | -3.27 | .001 |
| Condition × Bin. Out × Extremity | -0.06 | -2.02 | .044 |

*Note.* Logistic mixed-effects model, condition, the binary outcomes, the extremity of the feedback, and interaction terms as fixed effects, participants as random effects.

***Estimates***

At the end of the free choice phase, participants provided contingency estimates as well as conditional probability estimates. The pattern was similar for both measures, exhibiting a small bias in the reward-rich and no bias in the reward-impoverished condition.

Despite the inability of the reward-conditions manipulation to induce constant sampling preferences, the post-sequence estimates exhibited a slightly enhanced tendency to exploit (i.e., to favor the frequent location) in the reward-rich condition, which however was not quite significant in the Bayesian analysis ($M$ = 5.67, $SD$ = 31.52), $BF_{+0}$ = 1.02, $t(100)$ = 1.81, $p$ = .037, $d$ = 0.18, 95% $CI_d$ [-0.02, 0.38]. Estimates in the reward-impoverished condition did not reveal any sampling tendency ($M$ = -0.48, $SD$ = 30.16), $BF_{01}$ = 8.92, $t(99)$ = -0.16, $p$ = .874, $d$ = 0.02, 95% $CI_d$ [-0.21, 0.18]. However, as with our sampling measure, estimates did not differ between conditions ($BF_{+0}$ = 0.71, $t(198.77)$ = 1.41, $p$ = .079, $d$ = 0.20, 95% $CI_d$ [-0.08, 0.48]) nor was there an overall bias, $BF_{+0}$ = 0.28, $t(200)$ = 1.20, $p$ = .116, $d$ = 0.08, 95% $CI_d$ [-0.05, 0.22].

We also asked participants to indicate how likely they thought it was to win given that they chose either location. Participants' conditional probability estimates (frequent location: 57.34%, *SD* = 17.69; infrequent location: 51.90%, *SD* = 20.51) revealed no significant bias in the reward-rich condition, $\Delta P_{rich}$ = .05 (*SD* = 0.34), $BF_{+0}$ = 0.73, *t*(100) = 1.62, *p* = .054, *d* = 0.16, 95% $CI_d$ [-0.04, 0.36]. In the reward-impoverished condition (frequent location: 46.48%, *SD* = 22.56; infrequent location: 42.84%, *SD* = 19.33) on the other hand, we did find evidence in line with our hypotheses: The mean Δ*P*-score was $\Delta P_{impoverished}$ = .04 (*SD* = 0.33), which was more likely to represent the absence of any bias, $BF_{01}$ = 4.94, *t*(99) = 1.12, *p* = .267, *d* = 0.11, 95% $CI_d$ [-0.08, 0.31]. As on the other measures the two conditions did not differ from one another ($BF_{+0}$ = 0.21, *t*(198.89) = 0.38, *p* = .351, *d* = 0.05, 95% $CI_d$ [-0.22, 0.33]) and there was also no overall effect, $BF_{+0}$ = 0.97, *t*(200) = 1.95, *p* = .027, *d* = 0.14, 95% $CI_d$ [0.00, 0.28].

***Confidence***

Participants estimated the frequent location just as confidentially (44.26%, *SD* = 26.30) as they did the infrequent location (40.98%, *SD* = 24.84), $BF_{01}$ = 4.08, *t*(398.69) = 1.28, *p* = .200, *d* = 0.13, 95% $CI_d$ [-0.07, 0.32].

**Discussion**

In this experiment, we exposed participants to graded (not just dichotomous) outcomes, and without any initial estimates, participants were not obliged to stick to their initially induced, illusory preferences via self-consistency. Under these circumstances, we were unable to induce different sampling strategies that would serve to either maintain or correct for the initial biases. As a plausible consequence, then, our manipulation of reward-rich vs. reward-impoverished environments also failed to produce a significant difference in the final tendency to maintain or to correct for an initial bias, the effective induction of which was never checked. Without an initial judgment after the first stage, the pseudocontingency manipulation was apparently not effective.

Still, despite the apparent absence of a stable preference to continue sampling from the frequent location or to gather more information about the infrequent location, we did observe a strong and systematic influence of the hedonic reward manipulation on the tendency to shift or stay from trial to trial. While binding participants to initial contingency ratings in previous research may have induced a stable tendency to either exploit the frequent winner or else to explore for alternatives, our trial-by-trial analysis points to another systematic sampling effect, namely, win-stay-lose-shift. Although the shift rate (after losses) was generally higher in the reward-rich than in the reward-impoverished condition, this hedonic rule corroborates the psychological notion that reward-rich environments foster exploitation whereas

reward-impoverished settings encourage exploration. Because absolute loss rates are much lower in reward-rich settings (25% losses), there are more reasons to exploit and less reasons to explore than in reward- impoverished (75% losses) settings. Thus, fine-grained analyses at trial level complement analyses across the entire series in describing the impact of illusory inferences about a contingency.

**Experiment 2**

With respect to this hypothetical interpretation, Experiment 2 constitutes a replication of Experiment 1 with only one crucial difference: in line with earlier studies by Harris et al. (2020), we ask participants for estimates twice. First, after the initial evidence phase and before the free choice phase we ask them to indicate for the first time their relative contingency estimates, conditional probability estimates, and their confidence. Then again after sampling, we administer the same dependent variables as we did in Experiment 1 (i.e., we again collect the measures already asked in the premeasures as well as the measures for control and predictability). Thus, premeasures not only serve as a check on the effectiveness of our initial manipulation, but should also force participants to reflect on the experienced distribution. Granting a dominant pseudocontingency effect, (in)frequent outcomes should appear to be contingent on (in)frequent locations, producing a bias to favor frequent location in the reward-rich condition but in favor of the infrequent location in the reward-impoverished condition. In contrast, a dominant ratio bias should produce a preference for the frequent location in both reward conditions. A mixture of both biases, either in different participants or across choices within the same participants, should produce a general bias in favor of the frequent location, which should however be weaker in the reward-impoverished than in the reward-rich condition.

**Methods**

Participants were again recruited via the online crowdsourcing platform Prolific Academic. The experiment was run in English on SosciSurvey (Leiner, 2020). One hundred and ninety-nine participants ($N_{female}$ = 127) with an average age of 35 years (*SD* = 9.5) participated for a financial reward of at least £0.85 plus additional earnings (max £2.15, mean £1.50) based on performance. A sensitivity analysis using G*Power ($\alpha$ = 0.05, 1-$\beta$ = .80, *N* = 199) indicated that we could detect effect sizes as small as *d* = 0.177.

This experiment replicates Experiment 1 with one distinct change: After the initial evidence, but before participants started sampling freely, they indicated their relative contingency estimates, as well as conditional probability estimates and confidence therein.

During the initial evidence phase, participants were forced to select the same distribution as in Experiment 1 (see also Table 3.1), which we expected

to induce biased preferences. Depending on the condition, they would either experience predominantly positive or negative outcomes. Depending on the relative influence of pseudocontingency illusions and ratio biases, an exploitation tendency to sample predominantly from the frequent location should be either confined to or occur predominantly or maybe exclusive in the reward-rich condition compared to the reward-impoverished condition. However, any sampling bias induced in the initial phase should mediate the perceived contingency and the resulting preference after the free-sampling stage.



**Results**

***Estimates Pre-Measure***

After the induction phase, participants indicated their relative preference for either location as well as conditional probability estimates. In the reward-rich condition, with 9:3 positive and negative outcomes for a frequent location and 3:1 for an infrequent location, participants indeed provided more favorable ratings for the frequent than for the infrequent location. However, no reversal was obtained in the reward-impoverished condition; rather than favoring the infrequent location, participants did not form any initial biases at all. That is, they did not associate the infrequent location with the infrequent positive outcomes and the frequent location with the frequent negative outcomes. This pattern is clearly reflective of a mixture of both biases, the pseudocontingency and the ratio bias type.

More specifically, the relative contingency estimates indicated a preference for the frequent location not only in the reward-rich condition ($M = 14.69$, $SD = 26.86$, $BF_{+0} = 59{,}742.96$, $t(97) = 5.41$, $p < 0.001$, $d = 0.55$ , 95% $CI_d$ [0.33, 0.76]) but also in the reward-impoverished condition ($M = 8.24$, $SD = 28.77$), $BF_{-0} = 0.03$, $t(100) = 2.88$, $p$(one-tailed) = 0.998, $d = 0.29$ , 95% $CI_d$ [0.09, 0.49]. The difference between conditions did not reach a conventional level of statistical significance, $BF_{01} = 1.86$, $t(196.71) = 1.64$, $p = .103$, $d = 0.23$, 95% $CI_d$ [-0.05, 0.51]. Across conditions, we found a large overall preference for the frequent location, $BF_{+0} = 591{,}207.92$, $t(198) = 5.76$, $p < .001$, $d = 0.41$, 95% $CI_d$ [0.26, 0.55].

The conditional probability estimates in the reward-rich condition (frequent location: 65.32%, $SD = 18.41$; infrequent location: 48.78%, $SD = 18.54$) reflected a strong bias favoring the frequent location, $\Delta P_{rich} = .17$ ($SD = 0.31$), $BF_{+0} = 49{,}082.87$, $t(97) = 5.37$, $p < .001$, $d = 0.54$ , 95% $CI_d$ [0.33, 0.75]. In the reward-impoverished condition, however, participants did not exhibit a marked bias in favor of either location (frequent location: 45.04%, $SD = 22.08$; infrequent location: 40.10%, $SD = 22.81$), $\Delta P_{impoverished} = .05$ ($SD = 0.33$), $BF\text{-}_0 = 0.05$, $t(100) = 1.49$, $p = .930$ (one-tailed), $d = 0.15$ , 95% $CI_d$ [-0.05, 0.34]. A significant difference between conditions testified to biases deviating more strongly from chance level in the reward-rich as

opposed to the reward-impoverished condition ($BF_{01}$ = 0.31, $t$(196.33) = 2.56, $p$ = 0.011, $d$ = 0.36 , 95% $CI_d$ [0.08, 0.64]). Across conditions we again found an overall effect, $BF_{+0}$ = 3,699.25, $t$(198) = 4.63, $p$ < .001, $d$ = 0.33, 95% $CI_d$ [0.19, 0.47].

***Confidence Pre-Measure***

Subjective confidence tended to be slightly (but non-significantly) higher for estimates of the frequent (51.73%, $SD$ = 24.75) than the infrequent location (47.34%, $SD$ = 24.71), $BF_{01}$ = 1.99, $t$(396) = 1.77, $p$ = .077, $d$ = 0.18, 95% $CI_d$ [-0.02, 0.37].

***Sampling***

Figure 3.5 summarizes participants' choices across trials. We used the same logistic mixed-effects model as in Experiment 1 with trial number and condition as predictors and participants as random effects. A non-significant intercept again the lack of an initial bias in the reward-impoverished condition, c = 0.06, $z$ = 0.49, $p$ = .623. The positive beta-weight indicates that reward-impoverished participants also tended to favor the frequent location, even though it had been mainly paired with losses. It would seem that, in line with a ratio bias, the higher absolute number of winning opportunities for the frequent location dominated the sampling preferences. A significant main effect of condition indicates successful bias-induction in the reward-rich condition; participants sampled more from the frequent location, $\beta_1$ = 0.52, $z$ = 2.92, $p$ = .004. Neither trial number nor the interaction terms were significant (see full model in Table 3.4).

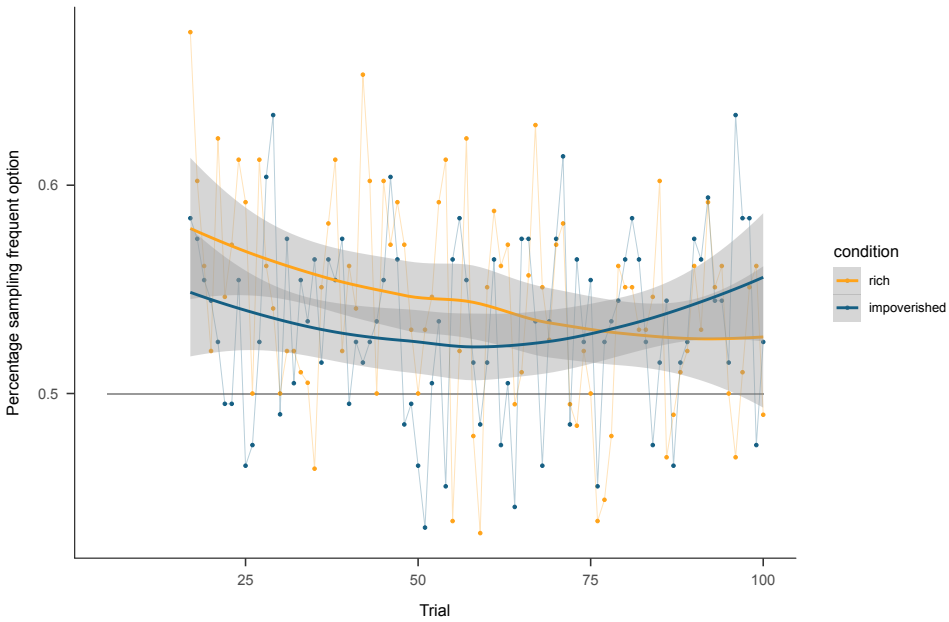



*Figure 3.5*. Percentage of participants sampling the frequent location per trial (Experiment 2). Includes also a local polynomial regression fit with 95% CI.

Table 3.4

*Regression table for choice model in Experiment 2*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| Intercept | 0.06 | 0.49 | .623 |
| Condition | 0.52 | 2.92 | .004 |
| Trial | -0.62 | -1.62 | .106 |
| $Trial^2$ | 0.81 | 1.77 | .076 |
| Trial × Condition | -0.13 | -0.24 | .813 |
| $Trial^2$ × Condition | -0.23 | -0.36 | .720 |

*Note.* Quadratic logistic mixed-effects model, trial number and condition as fixed effects, participants as random effects.



We then analyzed the tendency to shift or stay by means of the same logistic mixed effects linear model with participants as random effect and condition, the binary outcomes, as well as the extremity of the feedback as fixed effects. Figure 3.6 visually presents the means across participants; Table 3.5 contains the full mixed-effects models. A model including extremity once again outperformed a model without this factor, $\chi^2(4) = 111.19$, $p < .001$.

In an analysis with main effects only, the basic win-stay-lose-shift effect is again visible in a main effect for binary outcomes. Losses strongly increased the shift rate, $\beta = -0.39$, $z = -17.57$, $p < .001$. The main effects for condition and for extremity were insignificant (see Table 3.5).

Another analysis with interaction terms included yielded significant interactions between conditions and extremity, $\beta = -0.15$, $z = -4.71$, $p < .001$, and between binary outcomes and extremity, $\beta = -0.06$, $z = -2.80$, $p = .005$. Figure 3.6 shows that the unequal shift rate after losses and wins, diverging increasingly with extremity, was mainly due to the reward-rich condition, where the shift rate strongly decreased with extremity of wins but increased with extremity of losses. Neither the Condition × Binary outcome interaction nor the three-way Condition × Binary outcome × Extremity of feedback interaction were significant (see Table 3.5).

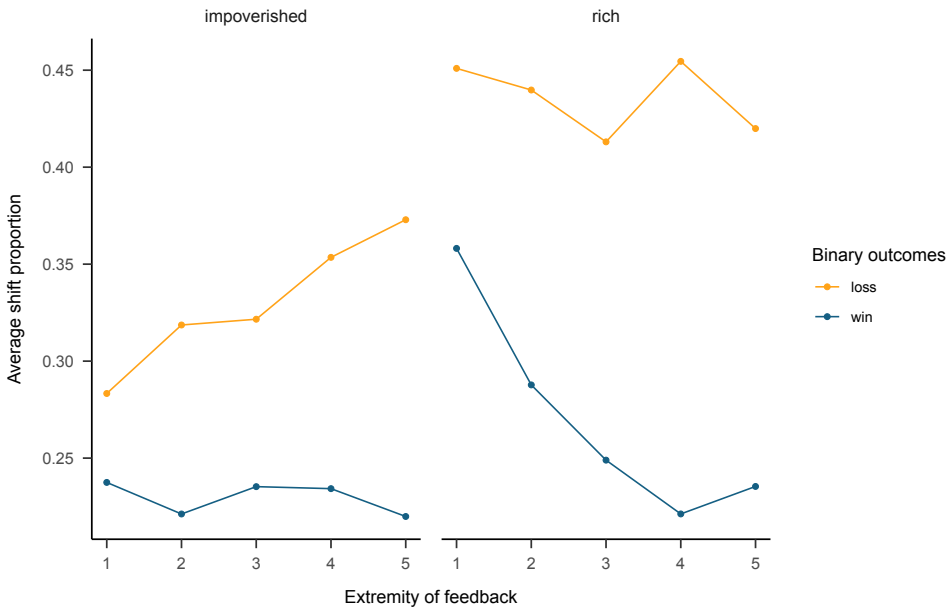



*Figure 3.6.* Average proportion of trials on which participants shift from previous option given the condition (reward-impoverished vs. reward-rich) and the extremity of feedback (Experiment 2).

Table 3.5

*Regression table for shift model in Experiment 2*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| **Main effects model** | | | |
| (Intercept) | -1.00 | -7.62 | < .001 |
| Condition | 0.00 | -0.02 | .983 |
| Bin. Out | -0.39 | -17.57 | < .001 |
| Extremity | -0.02 | -1.60 | .110 |
| **Interaction effects model** | | | |
| (Intercept) | -1.15 | -8.00 | < .001 |
| Condition | 0.41 | 2.04 | .041 |
| Bin. Out | -0.10 | -1.35 | .178 |
| Extremity | 0.05 | 2.17 | .030 |
| Condition × Bin. Out | -0.06 | -0.58 | .561 |
| Condition × Extremity | -0.15 | -4.71 | < .001 |
| Bin. Out × Extremity | -0.06 | -2.80 | .005 |
| Condition × Bin. Out × Extremity | -0.05 | -1.69 | .091 |

*Note.* Logistic mixed-effects model, condition, the binary outcomes, the extremity of the feedback, and interaction terms as fixed effects, participants as random effects.

***Estimates Post-Measure***

Following the free choice phase, we again asked participants to indicate their relative preference for either. Here, the same pattern we saw during sampling emerged even more strongly. Participants in the reward-rich condition exhibited a tendency to favor the frequent location, but so did participants in the reward-impoverished condition. Moreover, in their conditional probability estimates, participants again provided higher ratings for the frequent location in both the reward-rich and the reward-impoverished condition, suggesting an initial bias of the ratio-bias type rather than a pseudocontingency.

For the relative contingency estimates participants in the reward-rich condition indicated a (slight) preference for the frequent location ($M$ = 6.29, $SD$ = 28.98), $BF_{+0}$ = 1.97, $t(97)$ = 2.15, $p$ = .017, $d$ = 0.22, 95% $CI_d$ [0.02, 0.42]. But so did participants in the reward-impoverished condition ($M$ = 6.45, $SD$ = 26.42), $BF_{01}$ = 0.53, $t(100)$ = 2.45, $p$ = .016, $d$ = 0.24, 95% $CI_d$ [0.05, 0.44]. While conditions did not moderate these findings ($BF_{+0}$ = 0.15, $t(194.11)$ = -0.04, $p$ = .516, $d$ = -0.01, 95% $CI_d$

[-0.29, 0.27]), we did find an overall bias, $BF_{+0}$ = 25.26, $t$(198) = 3.25, $p$ < .001, $d$ = 0.23, 95% $CI_d$ [0.09, 0.37].

Also for the conditional probability estimates did participants show biases favoring the frequent location in both the reward-rich (frequent location: 61.77%, $SD$ = 20.14; infrequent location: 54.72%, $SD$ = 19.09, $\Delta P_{rich}$ = .07, $SD$ = 0.30, $BF_{+0}$ = 3.01, $t$(97) = 2.35, $p$ = .010, $d$ = 0.24, 95% $CI_d$ [0.04, 0.44]) and the reward-impoverished condition (frequent location: 43.79%, $SD$ = 21.26; infrequent location: 36.55%, $SD$ = 20.32), $\Delta P_{impoverished}$ = .07 ($SD$ = 0.26), $BF_{01}$ = 0.24, $t$(100) = 2.79, $p$ = .006, $d$ = 0.28, 95% $CI_d$ [0.08, 0.48]. Again conditions did not moderate this finding, $BF_{+0}$ = 0.15, $t$(192.12) = -0.05, $p$ = .52, $d$ = -0.01, 95% $CI_d$ [-0.29, 0.27], but we did find an overall effect, $BF_{+0}$ = 82.35, $t$(198) = 3.62, $p$ < .001, $d$ = 0.26, 95% $CI_d$ [0.12, 0.40].

***Confidence Post-Measure***

Participants estimated the frequent location just as confidentially (52.65%, $SD$ = 25.94) as they did the infrequent location (50.98%, $SD$ = 25.81), $BF_{01}$ = 7.37, $t$(395.99) = 0.65, $p$ = .519, $d$ = 0.06, 95% $CI_d$ [-0.13, 0.26].

**Discussion**

Thus, participants in both the reward-rich as well as the reward-impoverished condition displayed persisting biases favoring the frequent location. Initially, the strength of the induced bias seemed to differ between conditions. The initial conditional probability estimates seemed to indicate a bias toward the frequent location only for the reward-rich condition but not for the reward-impoverished condition. Such a one-sided initial bias is suggestive of a mixture of pseudocontingency and ratio biases. The final estimates, however, showed a general bias favoring the frequent location that was equally visible in both reward conditions. This slightly changing pattern could reflect the fact that contingency estimation tasks are relatively more sensitive to pseudocontingencies than hedonically motivated sampling and resulting preferences, which may by more prone to ratio biases. But in any case, our findings corroborate the notion that the persistence of an initial bias depends on the reward environment and the sampling strategy induced by whatever initial bias was at work. This holds regardless of whether the initial bias is of the pseudocontingency or of the ratio bias type.

The most reasonable explanation for the present experiment says that participants fell prey to a ratio bias (Denes-Raj et al., 1995; Reyna & Brainerd, 2008). As outlined in the introduction, a ratio bias amounts to disregarding the denominator (here the overall number of trials encountered with either location) and instead focusing solely on the wins. Then, participants in the reward-impoverished condition should associate the frequent location more strongly than the infrequent location with the opportunity of winning: While it also had proportionally more losses during the

initial evidence (nine), it did have more wins (three) than the infrequent alternative (three losses and one win).

In hindsight, the new evidence on win-stay-lose-shift conceived as exploration versus exploration at the trial level adds an additional layer of explanation why any pseudocontingency bias that might have been initially present (in the pre-estimates) cannot be expected to persist in the reward-impoverished condition. Although a pseudocontingency after 1 positive / 3 negative outcomes for the infrequent location and 3 positive / 9 negative outcomes for the frequent location should link losing with the frequent and winning with the infrequent location, any initial preference to sample from the infrequent location would not survive the high shift rate expected in a reward-impoverished setting with 75% losses. Thus, our new discovery of the win-stay-lose-shift regulation of item-level exploitation and exploration offers an intriguing explanation for the ineffectiveness of pseudocontingencies and the dominance of ration biases in the reward-impoverished conditions.



Closer analyses of the participants in the reward-impoverished condition indeed support the assumption that pseudocontingency (PC) biases were not fully missing; they were simply overruled by hedonic sampling rules and by the impact of a dominant ratio bias (RB). To further investigate this idea, we ran an extra analysis with a dummy variable to distinguish those $N_{PC}$ = 37 participants whose initial preference for the infrequent location reflected an initial pseudocontingency from those $N_{RB}$ = 64 participants whose preference for the frequent location indicated a ratio bias. With the reward-impoverished condition split up into these two groups, we indeed find that initial biases persist and influence the final contingency estimates even when the high shift-rate served to conceal the initial preferences during the free-sampling stage.

In an analysis of the sampling data, both groups hardly differed in their respective deviation from chance level, $BF_{+0}$ = 0.49, $t(89.55) = 0.94$, $p = .175$, $d$ = 0.18, 95% $CI_d$ [-0.23, 0.59]. Figure 3.7 reveals however that participants in the PC subgroup were initially more prone to sample from the infrequent group and only later converged with chance level, while such a convergence is less clear in the RB subgroup in which the initial biases apparently persisted and affected the final contingency judgments. In the analysis of the relative contingency measure, the deviance from chance level is weaker in the PC than in the RB subgroup, $BF_{+0}$ = 2.19, $t(79.48) = 1.98$, $p = .026$, $d = 0.40$, 95% $CI_d$ [-0.01, 0.81]. Separate analyses yield a significant bias favoring the frequent location in the RB subgroup ($BF_{+0}$ = 19.37, $t(63) = 3.08$, $p = .002$, $d = 0.39$, 95% $CI_d$ [0.13, 0.64]), but not in the PC subgroup, $BF_{01}$ = 5.65, $t(36) = -0.05$, $p = .963$, $d = -0.01$, 95% $CI_d$ [-0.33, 0.31].

Similarly, in the analysis of conditional estimates, we find an overall bias in the RB subgroup ($BF_{+0}$ = 24.45, $t$(63) = 3.17, $p$ = .001, $d$ = 0.40, 95% $CI_d$ [0.14, 0.65]), but not in the PC subgroup, $BF_{01}$ = 5.27, $t$(36) = 0.39, $p$ = .70, $d$ = 0.06, 95% $CI_d$ [-0.26, 0.39]. The difference between subgroups is only marginally significant in this analysis, $BF_{+0}$ = 1.40, $t$(80.05) = 1.71, $p$ = .045, $d$ = 0.35, 95% $CI_d$ [-0.07, 0.76].

Although these post-hoc analyses should be considered with caution, because sample size is low for these tests and the groups are unevenly distributed, the evidence is consistent with our general perspective on persistence of initial biases. Our twofold analysis of exploitation and exploration as a function of both environments (reward-rich vs. reward-impoverished) and trial outcomes (wins vs. losses) suggest at least a first idea of the underlying mechanism. To further consolidate our theoretical account, we conducted a third experiment, in which a new initial stimulus a distribution that should leave little doubt as to the initial bias and sampling preference.

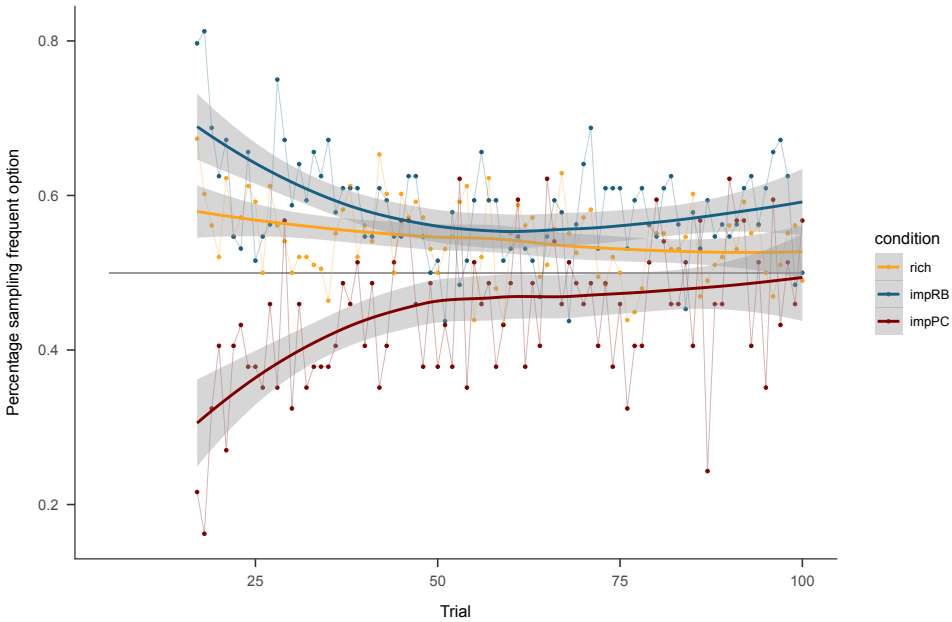



*Figure 3.7*. Percentage of participants sampling the frequent location pretrial for reward-rich and two reward-impoverished conditions (Experiment 2). Includes also a local polynomial regression fit (loess) with 95% CI.

## Experiment 3

Specifically, we used a distribution that implies a perfect contingency in the initial evidence. One location always wins, while the other always loses. With this distribution, we could expect uniform biases across participants due to a clear-cut contingency in the initial distribution. The aim of this experiment was to show in such an unequivocal setting that the intricacies encountered in the first two experiments do not reflect the non-viability of our theory but only competing inferences in a complex task setting. Assuming that using clear-cut contingencies to manipulate initial preferences can overrule such competition, we expected biases to maintain only in the reward-rich but to be attenuated in the reward-impoverished condition.



### Methods

Participants for this experiment were again recruited via the online crowdsourcing platform Prolific Academic and the experiment was run in English on SosciSurvey (Leiner, 2020). One hundred participants ($N_{female}$ = 59) with an average age of 32 years ($SD$ = 9.42) participated for a financial reward of £1.15 plus additional earnings (mean £1.40, max £1.60) based on performance. The sensitivity analysis using G*Power ($\alpha$ = 0.05, 1-$\beta$ = .80, $N$ = 100) indicated that we could detect effect sizes as small as $d$ = 0.250.

The induction phase lasted for four trials. Participants in the reward-rich condition won on three of the four trials after having to choose one location and lost on the one trial in which they had to choose the other location. The order was fixed so that they would win twice, lose once, and win one more time. For example, they would have to choose the street corner twice, winning both times by receiving between eight and twelve strawberry baskets. Then, they would have to choose the park and would lose (≤ 4 baskets), before having to sample the street corner one last time and winning. In the reward-impoverished condition, participants would lose three times and only win once on the third trial (Table 3.1).

### Results

#### *Sampling*

Figure 3.8 summarizes the sampling biases across trials. We again analyzed sampling using the same logistic mixed-effects model as in Experiments 1 and 2 with trial number and condition as predictors and participants as random effects (see also Table 3.6). The negative intercept indicates the successful induction of a bias towards the infrequent location in the reward-impoverished condition, $c$ = -0.24, $z$ = -1.84, $p$ = .066. The positive main effect of trial number ($\beta_2$ = 1.45, $z$ = 3.33, $p$ < .001) together with the negative quadratic main effect of trial number ($\beta_3$ = -1.38, $z$ = -3.07, $p$ = .002) indicate an inverse u-shaped trend with increasing trial number seeing first a shift towards chance level but then a later decline back towards the initial bias.

The positive main effect for condition indicates a bias towards the frequent location in the reward-rich condition, $\beta_1 = 0.79$, $z = 4.32$, $p < .001$. This main effect is qualified by interactions both with trial number ($\beta_4 = -2.06$, $z = -3.32$, $p < .001$) and with the quadratic main effect of trial number ($\beta_5 = 1.55$, $z = 2.43$, $p = .015$). The difference between the reward-rich and the reward-impoverished condition decreases with increasing trial number but later increases again. Figure 3.8 shows that this is mainly due to the reward-impoverished condition.

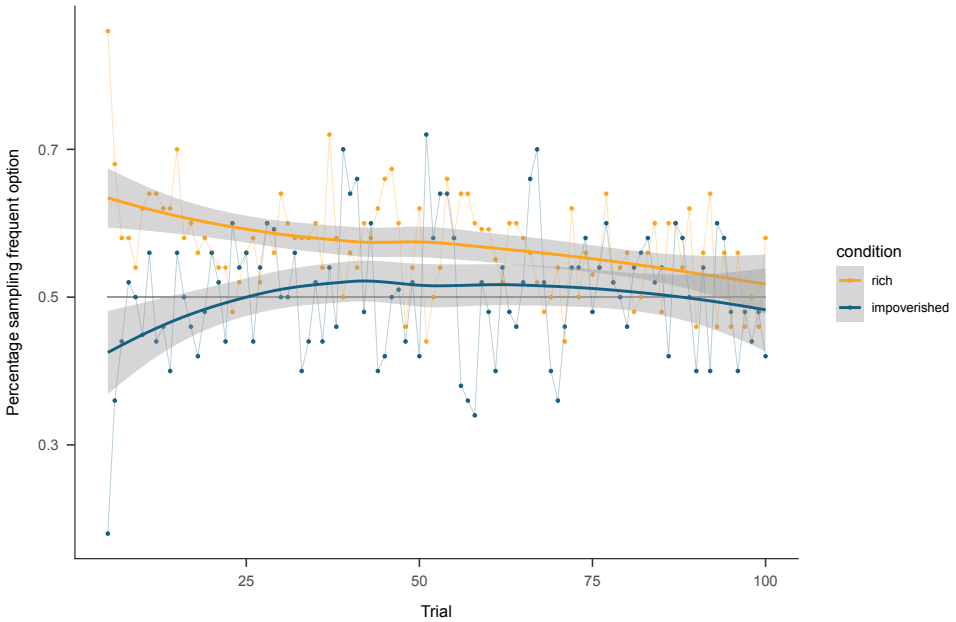



*Figure 3.8*. Percentage of participants sampling the frequent location per trial (Experiment 3). Includes also a local polynomial regression fit with 95% CI.

Table 3.6

*Regression table for choice model in Experiment 3*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| Intercept | -0.24 | -1.84 | .066 |
| Condition | 0.79 | 4.32 | < .001 |
| Trial | 1.45 | 3.33 | < .001 |
| $Trial^2$ | -1.38 | -3.07 | .002 |
| Trial × Condition | -2.06 | -3.32 | < .001 |
| $Trial^2$ × Condition | 1.55 | 2.43 | .015 |

*Note.* Quadratic logistic mixed-effects model, trial number and condition as fixed effects, participants as random effects.



We once again used the same logistic mixed effects linear model with participants as random effect and condition, the binary outcomes, as well as the extremity of the feedback as fixed effects. Figure 3.9 visually presents the means across participants; Table 3.7 contains the full mixed-effects model. A model including extremity once again performed better than a model without this factor, $\chi^2(4) = 121.96$, $p < .001$.

In an analysis with only main effects, we again found the main effect for binary outcomes such that losing increased the probability of shifting, $\beta = -0.48$, $z = -16.81$, $p < .001$. However, we now also found a main effect for condition such that participants in the reward-rich condition tended to shift more so than participants in the reward-impoverished condition, $\beta = 0.47$, $z = 2.30$, $p = .022$. Extremity of the feedback was not significant in this experiment (see Table 3.7).

Including interaction terms resulted in two significant interaction results. We found the same pattern as in Experiment 2. Specifically, a Condition × Extremity interaction shows that the differential shift rate after wins and losses increased more strongly with increasing extremity in the reward-rich compared to the reward-impoverished condition (see Figure 3.9), $\beta = -0.18$, $z = -4.42$, $p < .001$. And we again found the interaction between binary outcomes and the extremity of feedback such that more extreme feedback resulted in stronger accentuation between wins and losses, $\beta = -0.12$, $z = -4.16$, $p < .001$. Neither the interaction effect for condition and binary outcomes nor the three-way interaction were significant (see Table 3.7).

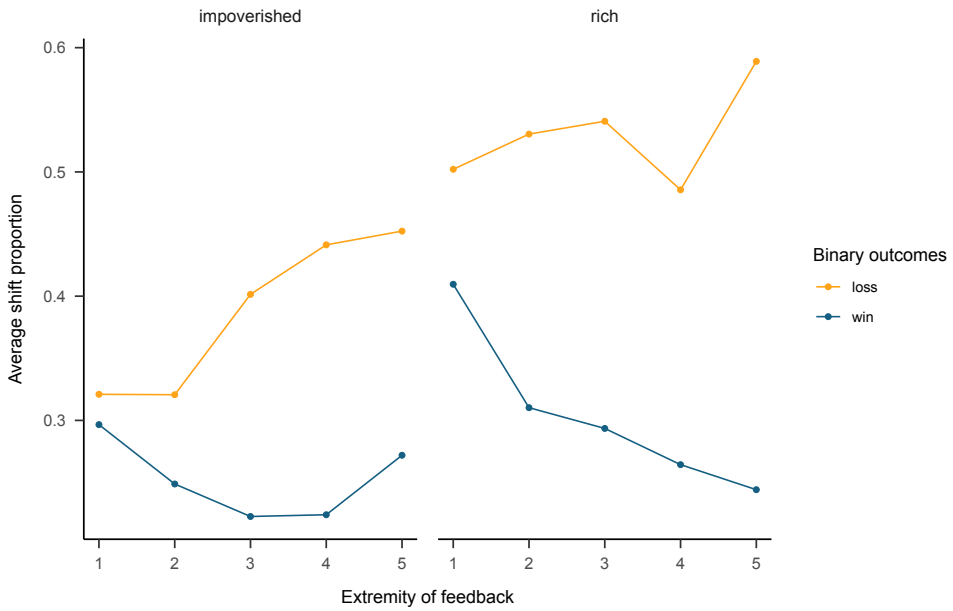



*Figure 3.9.* Average proportion of trials on which participants shift from previous option given the condition (reward-impoverished vs. reward-rich) and the extremity of feedback (Experiment 3).

Table 3.7

*Regression table for Experiment 3: shift model*

| Effect | Beta | z-value | p-value |
|---|---|---|---|
| **Main effects model** | | | |
| (Intercept) | -0.99 | -6.51 | < .001 |
| Condition | 0.47 | 2.30 | .022 |
| Bin. out | -0.48 | -16.81 | < .001 |
| Extremity | 0.01 | 0.58 | .561 |
| **Interaction effects model** | | | |
| (Intercept) | -1.13 | -6.62 | < .001 |
| Condition | 0.96 | 4.00 | < .001 |
| Bin. out | 0.01 | 0.16 | .874 |
| Extremity | 0.08 | 2.84 | .005 |
| Condition × Bin. out | -0.20 | -1.52 | .128 |
| Condition × Extremity | -0.18 | -4.42 | < .001 |
| Bin. out × Extremity | -0.12 | -4.16 | < .001 |
| Condition × Bin. out × Extremity | -0.03 | -0.68 | .494 |

*Note.* Logistic mixed-effects model, condition, the binary outcomes, the extremity of the feedback, and interaction terms as fixed effects, participants as random effects.

***Estimates***

We found this same pattern after the free choice phase when we asked participants to indicate relative contingency estimates. The data suggests that participants in the reward-rich condition had maintained their bias towards the frequent location. Participants in the reward-impoverished condition, on the other hand, had attenuated the initial bias.

Participants' conditional probability estimates again showed in the same pattern. In the reward-rich condition, estimates reflected a maintained bias. In the reward-impoverished condition, the initial bias was apparently attenuated.

An analysis of the relative contingency estimates suggests that participants in the reward-rich condition had maintained their bias towards the frequent location ($M$ = 14.44, $SD$ = 28.50), $BF_{+0}$ = 71.00, $t(49)$ = 3.58, $p$ < .001, $d$ = 0.51, 95% $CI_d$ [0.21, 0.80]. Participants in the reward-impoverished condition, on the other hand, had attenuated the initial bias ($M$ = -4.04, $SD$ = 30.59), $BF_{01}$ = 4.31, $t(49)$ = -0.93, $p$ = .355, $d$ = -0.13, 95% $CI_d$ [-0.41, 0.15]. We found support for a difference in the strength of the effect between the two conditions ($BF_{+0}$ = 1.57, $t(97.51)$ = 1.76, $p$ =

.041, $d$ = 0.35, 95% $CI_d$ [-0.05, 0.75]) as well as support for an overall bias, $BF_{+0}$ = 0.82, $t(99)$ = 1.69, $p$ = .047, $d$ = 0.17, 95% $CI_d$ [-0.03, 0.37], though the Bayesian analyses were inconclusive.

Analyses of conditional probability estimates exhibited the same pattern. In the reward-rich condition, participants' estimates (frequent location: 63.60%, $SD$ = 21.00; infrequent location: 49.38%, $SD$ = 21.05) indicated a maintained bias, $\Delta P_{rich}$ = .14 ($SD$ = 0.33), $BF_{+0}$ = 16.25, $t(49)$ = 3.01, $p$ = .002, $d$ = 0.43, 95% $CI_d$ [0.14, 0.72]. In the reward-impoverished condition (frequent location: 49.80%, $SD$ = 25.71; infrequent location: 56.66%, $SD$ = 24.51), the initial biases were again attenuated. The mean $\Delta P$-score was $\Delta P_{impoverished}$ = -.07 ($SD$ = 0.30), suggesting absence of any bias, $BF_{01}$ = 1.97, $t(49)$ = -1.60, $p$ = .115, $d$ = -0.23, 95% $CI_d$ [-0.51, 0.05]. However, the contrast between conditions was not statistically significant ($BF_{+0}$ = 0.66, $t(97.05)$ = 1.16, $p$ = .125, $d$ = 0.23, 95% $CI_d$ [-0.17, 0.63]) nor was there support for an overall effect, $BF_{+0}$ = 0.34, $t(99)$ = 1.10, $p$ = .137, $d$ = 0.11, 95% $CI_d$ [-0.09, 0.31].

***Confidence***

Participants were just as confident in estimating the frequent (54.50%, $SD$ = 28.21) as they were in estimating the infrequent location (54.10%, $SD$ = 27.14), $BF_{01}$ = 6.47, $t(197.7)$ = 0.10, $p$ = .919, $d$ = 0.01, 95% $CI_d$ [-0.26, 0.29].

**Discussion**

In this third experiment, the initial evidence so clearly favored one location over the alternative that we do not expect competing heuristic processes to have strongly influenced the bias elicited. And indeed, we found the pattern we predicted, with participants in the reward-rich condition displaying strong biases throughout the experiment while participants in the reward-impoverished condition quickly attenuated their initial biases. Taken together with the previous experiments, and in particular the post-hoc analyses of Experiment 2, these results add empirical support to the viability of our theory. When competing processes in the bias induction phase are removed, participants follow precisely the patterns we predicted, and that previous research has also found: As little as four trials suffice to seduce participants to exploit a seemingly better location which results in persisting biases in a reward-rich but not in reward-impoverished environment.

## General Discussion

In the current paper we investigated how the motivation to seek out hedonically positive outcomes can lead to persisting biases. That is, we investigated how initial biases can be maintained despite continuous interaction with one's environment and therefore supposedly ample opportunity to overcome such initial biases. More specifically, we tested the boundary conditions of effects demonstrated in earlier research (Harris et al., 2020). In these studies, it was found that inbalances in initial

forced sampling of two choice options resulted in percieved pseudocontingecies that biased sampling towards the frequently presented option. This bias was found to be mainained during subsequent free sampling.

Testing the limits of this effect in the current studies, we obtained intruiging patterns that sometimes deviated substantially from earlier findings (Harris et al., 2020). In contrast to earlier studies, participants knew from the outset which outcomes were positive and which were negative. Furthermore, these outcomes were not dichotomous, but graded. Finally, we did not always ask participants to indicate estimates immediately following the induction phase which could lead to self-consistency restraints and biases due to processes other than those of interest here. These changes generated conditions that were less artificial than in previous research, potentially adding to the generalizability of the findings.



Although in some respects the pattern of results differed substantially from previous research, one important finding emerged across all experiments: Once participants had completed the initial forced sampling phase, biases were more persistent over time when the environment offered many rewards, encouraging exploitation. The post-hoc analyses in Experiment 2 demonstrate this nicely. Participants that had an initial bias of the pseudocontingency type maintained this bias in reward-rich but attenuated this bias in reward-impoverished environments. In the latter, frequent negative outcomes dissuaded exploitation of any one option. Moreover, Experiment 3 revealed that a clear manipulation of actual contingencies in the forced sampling phase led to a maintained bias in the reward-rich, but not the reward-impoverished condition. In this sense, the current findings attest to the robustness of the findings reported in (Harris et al., 2020) and demonstrate their generalizability to situations with graded outcomes.

However, the biases that emerged as a result of our forced sampling manipulation sometimes generated surprising findings. In Experiment 2, the results also suggest that some participants developed a ratio bias, instead of a pseudocontingency bias. This could be the result of the rewards being present in the forced sampling phase, which directed participants' attention mainly to rewarding outcomes.

Through our trial-by-trial level analyses this paper also makes a second important contribution. They illustrate that even when global trends of exploration or exploitation cannot be found, those might still readily take place at the trial level in the form of a strategy of win-stay-lose-shift. Frequent wins sensitized participants in the reward-rich condition to losses and led to a strong trend of switching following a loss but staying following a win. These trends were far less pronounced in the reward-impoverished condition, in which frequent losses made any strategy seem little

effective. This, too, can be indicative of exploitative behavior: An option is exploited until it results in a loss, then quickly abandoned in favor of the next best alternative. Our findings suggest that exploitation can take place at the local and at the global level and that it is dependent on the characteristics of the environment. Even when complex task environments that allow for multiple competing processes overshadow any global exploitation trends, trial-by-trial behavior still readily displayed hedonically motivated behavior.

***Implications***

Why do people at times maintain unrealistic and strong beliefs regarding the outcomes of their actions even in situations in which they can continuously interact with their environment and in which they should be motivated to learn the best choice alternative? Why do some people believe so strongly in some alternative medicines for which medical evidence is far less certain than people's beliefs might suggest? We believe this to be an interaction between the information decision makers draw from reward-rich or reward-impoverished environments and premature exploitation based on (misleading) initial evidence. Goal-directed behavior oftentimes encourages exploitation more so than extended exploration phases. As a result, initial evidence can easily seduce people into engaging prematurely with a supposedly best option. In reward-rich environments, the information they draw supports their previous decisions and so they are continuously reaffirmed in their strategy. If a consumer has the choice between multiple brands of a new product and most brand alternatives are good, they might readily settle on any one option they happened to have an initial good experience with even when other alternatives might have been even better. Or, to come back to the medical example, when one's health is likely to improve regardless of whether and what medicine one takes, one might readily maintain the belief that a particular choice is the best option possible. To be sure, other (cognitive and motivational) processes most certainly also take place. Motivated reasoning, such as people upholding their views of themselves (Tajfel & Turner, 1979) might accentuate or overrule the processes described here. Nonetheless, we demonstrate that even when people encounter frequent opportunities to put their beliefs to the test and when they should clearly be motivated to exploit the best option, persisting biases can survive extended sampling periods. At best this is of little consequence or only leads to an unwarranted, idiosyncratic preference for a particular brand option. But at worst, it leads to people taking inferior medicine with detrimental costs to the health care system and potentially even one's individual health.

# CHAPTER 4

# Chapter 4:
## Is Knowledge Key? On the Moderating Role of Initially Knowing the Valence of Outcomes in Experience Sampling

**This chapter is based on:**

Harris, C., & Custers, R. (2021). Is knowledge key? On the moderating role of initially knowing the valence of outcomes in experience sampling. *Unpublished manuscript.*

## Abstract

Recent work (Harris, Fiedler, et al., 2022; Harris et al., 2020) has demonstrated that initial biases can be maintained even when participants undergo extended further sampling. They furthermore suggested that different initial biases can lead to the maintenance of biases under different settings and discussed initial knowledge of the valence of all outcomes as an important moderator for which type of bias participants would form initially. Here, we manipulate this potential moderator but find that knowledge over the valence does not seem to be a key moderator in explaining when participants form what type of cognitive bias.

One question that decision makers have to ask themselves in repeated choice tasks is at what point they believe to have found the best alternative. Should they continue exploring different alternatives or should they exploit a supposedly best option thereby maximizing outcomes? Recent work suggests that under certain conditions, decision makers tend to transition to a strategy of exploitation too readily which in turn can lead to unwarranted beliefs about choice alternatives being maintained (Harris, Fiedler, et al., 2022; Harris et al., 2020). This maintenance of unwarranted beliefs takes place even during extended sampling and even when decision makers should be motivated to learn and use the best choice alternative. In the above-mentioned experiments, the authors rely on pseudocontingencies (Fiedler et al., 2009) for inducing biases. However, the data suggests that at times another bias, denominator neglect (Reyna & Brainerd, 2008), was induced which in turn affected participants' behavior and consequentially the maintenance or attenuation of an initial bias differently than pseudocontingencies did. One possible moderator in explaining the induction of either a pseudocontingency bias or a denominator neglect bias might have been initial valence about the hedonic value of the outcomes. In the research reported in this article, we investigate this potential influence more carefully by explicitly manipulating whether participants have knowledge about the hedonic valence of outcomes at the outset of the task.



**Pseudocontingencies and Denominator Neglect**

Pseudocontingencies are a cognitive algorithm that rely on the alignment of skewed base rates in the inference process (Fiedler & Freytag, 2004). They have been shown to be a robust phenomenon in differences between social groups (Kutzner & Fiedler, 2017; Meiser & Hewstone, 2004), teachers' ratings of students (Fiedler et al., 2007), brand ratings of products (Vogel & Kutzner, 2017), or lottery tasks (Meiser et al., 2018). Harris et al. (2020) also found pseudocontingency effects and demonstrated how such a bias could be maintained throughout an extended sampling phase: If the distribution of initial evidence invited pseudocontingency-type inferences, participants relied on these inferences too readily. Instead of exploring alternatives more thoroughly, they exploited this seemingly better option. In environments in which most outcomes were positive, this led to the maintenance of the initial pseudocontingency bias while in environments in which most outcomes were negative, exploitation was not attractive and the initial pseudocontingency bias was attenuated as participants then explored alternatives. These findings speak to the importance of pseudocontingency inferences: Not only do they affect cognitive inferences and decisions in the moment, but their influence can be long-lasting and influence behavior over extended periods of time (Harris, Fiedler, et al., 2022; Harris et al., 2020; Meiser et al., 2018).

Nonetheless, at least in one experiment, Harris, Fiedler, et al. (2022) found effects indicative not of a pseudocontingency effect, but of denominator neglect. Denominator neglect occurs when the denominator of a ratio is ignored or underweighted, leading to people estimating 10/200 to be more likely than 1/20 (Denes-Raj et al., 1995; Reyna & Brainerd, 2008). To clarify, the distribution of initial evidence in the experiment by Harris, Fiedler, et al. (2022) had participants see one option lose nine times and win three times and the other option lose three times and win one time in a reward-impoverished condition. One option, option A, was thus more frequently shown than the alternative, option B. And one outcome, losses, was more frequent than the alternative, wins. A pseudocontingency would then predict base rate alignment such that the frequent option would be associated with the frequent outcome, and the infrequent option would be associated with the infrequent outcome. Option A should then be associated more strongly with losing and option B more strongly with winning making the latter the preferential option. Denominator neglect, on the other hand, would predict the neglect of the total number of times either option has been seen. Instead, the focus would be solely on the outcome of interest. Thus, instead of considering the three wins seen for option A relative to the nine losses, the losses would be neglected. Then, option A would see preferential compared to option B, as the first was seen winning more often than the latter disregarding the total number either option was seen. In a reward-rich condition, however, in which wins are more frequent than losses (e.g., option A wins nine times and loses three times, option B wins three times and loses one time), both pseudocontingencies and denominator neglect would predict preference for the same option.

But not only would a pseudocontingency and denominator neglect predict differing initial biases to arise in reward-impoverished conditions. As sampling then continues, further exploitation based on pseudocontingencies would quickly lead to losses with the supposedly better option which in turn would dissuade exploitation and thus result in the attenuation of initial biases. Exploitation based on denominator neglect, on the other hand, would result in the exact same pattern encountered so far: Many losses (that are neglected), but at the same time a higher number of positive outcomes with the frequent compared to the infrequent option. In this scenario we would predict the maintenance of initial biases.

There are a number of differences between Harris, Fiedler, et al. (2022) and typical experiments from the pseudocontingency literature, including Harris et al. (2020). In short, where categories are typically naturally distinct in the pseudocontingency literature (e.g., wins vs. losses or social groups), the experiments in Harris, Fiedler, et al. (2022) relied on a continuous scale that was artificially divided into positive and negative outcomes. Additionally, outcomes were graded

thereby further ambiguating the clear differentiation into groups that would invite the consideration of base rates in the inference process. Another interesting difference was that the valence of outcomes was initially known in this experiment while in Harris et al. (2020) that was not the case. This would, of course, have a strong influence on what information decision makers pay attention to: Are they forced to consider all outcomes as they are still unaware of whether certain outcomes are of particular interest? Or do they selectively focus only on certain outcomes? Answering these questions would not only help understand the results found by Harris, Fiedler, et al. (2022), but they would also be of interest to the wider pseudocontingency literature.

Knowledge about the valence of outcomes might be an important moderator in predicting the occurrence of pseudocontingencies or other cognitive illusions. Not only would a decision maker not know the valence of outcomes in a completely new situation, but this can also occur in only semi-familiar situations. A decision-maker walking through a town that they are not highly familiar with might only after a while decide that certain cues guiding them to public restroom facilities suddenly have a high personal hedonic value. While they might have initially paid attention to all signs, now one type of sign is of particular interest. They base their decision on the valence with all signs encountered in mind. Imagine, however, a decision maker that as they leave the parking lot knows they will want to find public restrooms quickly. They will likely blend out other signs and instead focus mainly on signs for public restroom facilities. More generally speaking, decision makers oftentimes may be focusing their attention on hedonically positive outcomes, thereby neglecting other outcomes in their inferences. And indeed, some have suggested that positive information is integrated and updated far more readily than negative information (Sharot et al., 2011; Sharot & Sunstein, 2020), but see also Harris and Hahn (2011). In conclusion then, knowledge about the valence of outcomes may be an important moderator in the type of cognitive illusion decision makers form, which in turn may strongly influence whether biases may be maintained or attenuated. As such this is not only an interesting boundary condition to the pseudocontingency literature, but of strong relevance to the decision-making field as a whole.

**The Present Study**

In order to investigate this potential moderator more closely, we conceptually replicated the strawberry paradigm used by Harris, Fiedler, et al. (2022). Some changes were necessary, however, so that we could manipulate knowledge of the valence of outcomes. In the original paradigm, participants could choose either of two locations to sell strawberries from. Outcomes were symbolically represented by strawberry baskets with a break-even point defined at six baskets. All outcomes below (ranging from 0 – 4) were to be considered losses and all outcomes above

(ranging from 8 – 12) were to be considered wins. It was therefore always clear to participants what the positive outcomes were. Here, we introduce a second fruit so that participants sold strawberries and grapes. The total amount of fruit was always identical (twelve symbols), so that participants sold either a lot of grapes and few strawberries or a few grapes and a lot of strawberries. Only one of these two fruits would later earn them points (and thus money), while the other could be ignored. We manipulated whether participants knew which fruit would earn or lose them points before encountering a distribution of initial evidence or whether they learned this only after encountering the initial evidence. Knowing the task valence might influence whether decision makers pay attention to selective information or the overall base rates and this in turn might have a strong influence on the mental processing style that could result in a cognitive illusion being formed that is either in the form of pseudocontingencies or denominator neglect. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the experiments (Simmons et al., 2012). Our preregistration can be found at https://osf.io/rk748/?view_only=11c319a8d1594e3688cb15a8de1b87e3.

## Methods

Participants for this experiment were recruited via the online crowdsourcing platform Prolific Academic and the experiment was run in English on SosciSurvey (Leiner, 2020). One hundred and ninety-eight participants ($N_{female}$ = 91) with an average age of 29 years (*SD* = 9.23) participated for a financial reward of at least £1.25 (max £2.05, mean £1.65) based on performance.

### Design

The experimental design is similar to Harris, Fiedler, et al. (2022): Participants engaged in a two-armed bandit task that consisted of 100 trials total in which they could choose between two locations, a street corner and a park, to sell fruit from. Selecting either location (Figure 4.1) would result in strawberries and grapes being sold. Only one of the two fruits would earn or lose participants points while the other fruit had no effect whatsoever. Depending on the condition participants learned about the payoffs of the fruits sooner or later in the task.

The overall amount of fruit sold, depicted symbolically by grape bundles and strawberry boxes, was always the same. However, participants would find themselves either selling many grapes and few strawberries, or the opposite. The fruit that affected payoffs was always the infrequent fruit that resulted in earning points on only 25% of trials. So, for example, while participants always sell 12 symbols representing the two fruits, they would sell a lot of grapes and few strawberries on 75% of trials and few grapes and many strawberries on 25% of trials. But they would only earn or lose points depending on how many strawberries they sold while

the number of grapes had no effect. As such, all participants were in a reward-impoverished condition in which they often lost and only seldom won.

Crucially, we manipulated when participants learned about the payoffs for the fruits. In the rewards-known condition, participants learned which fruit would earn or lose them points and which fruit had no influence before encountering the initial evidence. In the rewards-unknown condition, participants only learned this information after encountering the initial evidence, but before the first dependent variables (DVs) were answered. We counterbalanced which location was shown more often in the initial evidence phase as well as the orientation of the feedback (see also Figure 4.2).



*Figure 4.1*. The two locations participants could choose: a street corner and a park.

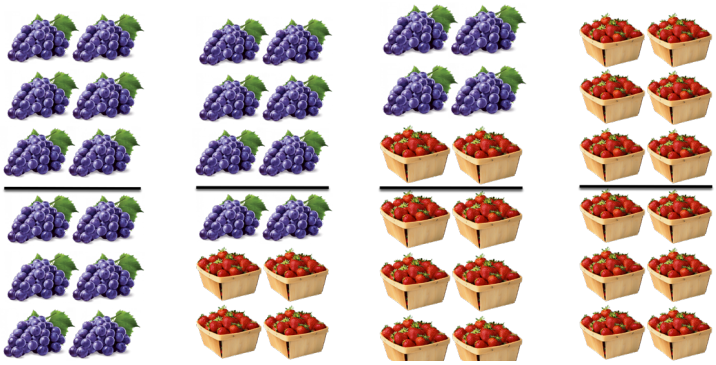


*Figure 4.2*. Example outcomes for trials with losses ranging from 0 – 4 baskets and wins from 8 – 12 baskets.

**Procedure**

Participants were introduced to the task in which they were to sell fruit from either of two locations, a street corner or a park. They were told that they would sell the same overall amount of fruit on any given trial, but that the ratio between strawberries and grapes would vary. In the rewards-known condition participants were then informed that they would only make money by selling one of the two

fruits. For example, they were informed that they could earn points when selling many strawberries and would lose points when selling few strawberries but that the number of grapes sold would not affect their payoffs.

All participants were then informed that in order to get to know the task better the computer had chosen a random distribution for the first few trials. This initial evidence was the distribution of 16 trials shown in Table 4.1, the trials of which were presented in random order. On each trial, after selecting a location, the outcome would be presented by means of the images as shown in Figure 4.2, as well as text ("You sold from location 'street' ['park'] and sold many [few] strawberries"). After a short delay the next trial started.

Participants in the reward-unknown condition were informed of the which fruits would earn or lose them points only after presentation of the initial evidence. They received the exact same information as the rewards-known condition. However, at that point they had already encountered the initial evidence.

Then, all participants were asked to answer three DVs in which they indicated relative contingency estimates and conditional probability estimates for the two locations as well as confidence estimates for their conditional probability estimates. For the relative contingency estimate, we asked participants from which of the two locations they were more likely to have sold many strawberry boxes (grape bundles). For the conditional probability estimates, we asked participants how likely they were to sell strawberry boxes (grape bundles) given that they had selected the street corner and the park, respectively. And for the confidence estimates, we asked participants how confident they were that they could make a reasonable estimate for the two conditional probability estimates. All three measures were answered on sliders with the two locations as the extremes for the relative contingency estimate and 0% and 100% as the extremes for both the conditional probability and the confidence estimates.

After answering these DVs, participants were reminded of their goal to earn as many points as possible during the experiment. Then, for the remaining 84 trials, they could choose freely on each trial which location to select. With the conclusion of this free sampling phase, we asked participants the same DVs as before, asked some basic demographic data, and concluded the experiment.

Table 4.1

*Distributions of initial evidence*

| | Experiment 1 | | |
|---|---|---|---|
| Reward-impoverished | Wins | Losses | |
| Frequently shown location | 3 | 9 | 25% |
| Infrequently shown location | 1 | 3 | 25% |
| | | | $\Delta P = 0$ [no contingency] |

*Note*: The distributions used for the initial evidence in the induction phase. The percentages depict the ratio of wins out of all trials for the respective location. $\Delta P$ is the difference score between the conditional probabilities and describes the contingency between location and outcome (Allan, 1980)



**Data Preparation**

All data preparation and analyses were undertaken with R (R Core Team, 2018) and in particular the packages 'dplyr' (Wickham et al., 2019) and 'BayesFactor' (Morey & Rouder, 2018). We recoded all measures so that positive values would indicate participants forming pseudocontingency inferences, while negative values would indicate participants forming inferences in line with denominator neglect. We expected participants in the rewards-unknown condition to initially use base rate alignment in the form of pseudocontingencies to infer the supposedly preferable alternative. In our coding estimates in line with this hypothesis would be indicated by estimates above chance level. We expected participants in the rewards-known condition to ignore the fruit that was of no consequence and focus solely on the wins with the fruit that affected payoffs. We assumed these participants would exhibit denominator neglect in their inference process, which in our coding would result in estimates below chance level.

We expected opposite biases in the rewards-known and the rewards-unknown condition and performed one-sided tests ($BF_{+0}$ and $BF_{-0}$) for the first measurement point. We expected these initial biases to be attenuated during sampling in the rewards-unknown condition and tested for this with two-sided tests ($BF_{01}$) for the second measurement point. In the rewards-known condition, on the

other hand, we expected these initial biases to be maintained and again performed one-sided tests. All Bayesian tests use the default settings of the BayesFactor package, including a Cauchy distribution of width $r$ = . In appendix C, 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. In all graphs the measure of dispersion are confidence intervals, which we also report for the effect sizes, denoted with a subscript d ($CI_d$). The analyses and data can be found on an Open Science Framework repository (https://osf.io/kedgp/?view_only=256ca3a2db2948e28a4632ee038cdad7).

## Results

### Relative Contingency Estimate Pre-Measure

After the induction phase and after participants in the rewards-unknown condition had also learned which fruit would affect the final payoffs and which fruit would not, we asked participants to indicate relative contingency estimates regarding either location. Surprisingly, participants' estimates in the rewards-unknown condition indicated no bias for either option ($M$ = -0.05, $SD$ = 33.30), $BF_{+0}$ = 0.11, $t$(99) = -0.02, $p$ = .506, $d$ = 0.00, 95% $CI_d$ [-0.20, 0.19]. Participants in the rewards-known condition did indicate a bias, however in the opposite direction of what we hypothesized ($M$ = 3.77, $SD$ = 30.16), $BF_{-0}$ = 0.05, $t$(97) = 1.24, $p$ = .89, $d$ = 0.12, 95% $CI_d$ [-0.07, 0.32]. See also Figure 4.3.

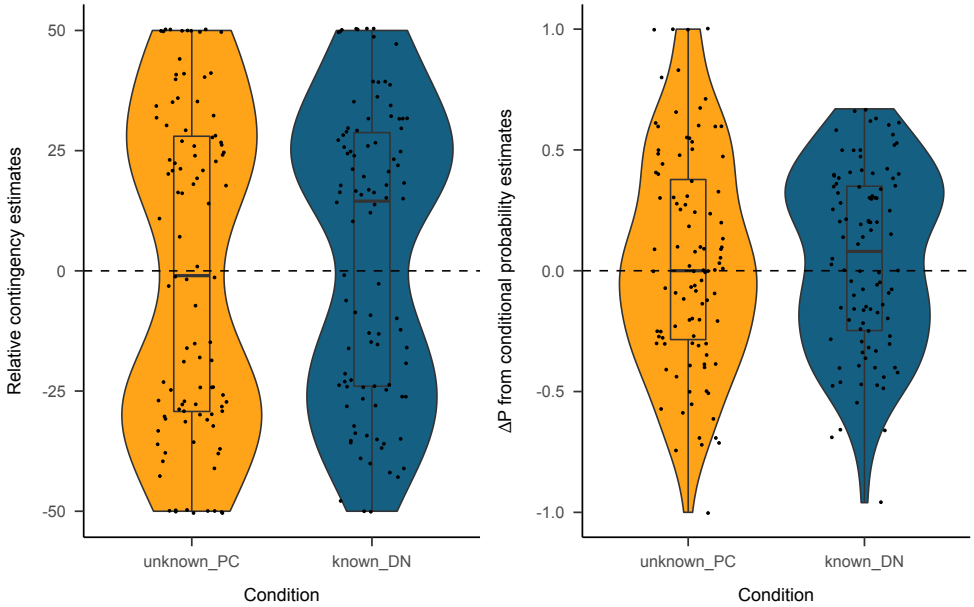



*Figure 4.3.* Estimates for premeasures of relative contingency and conditional probability estimates.

**Conditional Probability Estimates Pre-Measure**

We then asked participants to indicate how likely it was to sell many of whichever fruit affected payoffs. From these estimates we calculated the difference between the two conditional estimates, $\Delta P$ (Allan, 1980), which describes the estimated contingency between locations and selling a lot of the target fruit. In the rewards-unknown condition, participants estimated the probability to sell a lot of the target fruit at 50.37% ($SD$ = 26.04) for the infrequent and at 46.13% ($SD$ = 25.63) for the frequent location. The contingency estimate did only descriptively supported our hypothesis with the data hinting that the infrequent location was associated more strongly with winning than the frequent location though this was not supported by inferential tests ($\Delta P = 0.04$, $SD = 0.45$), $BF_{+0} = 0.28$, $t(99) = 0.95$, $p = .173$, $d = 0.09$, 95% $CI_d$ [-0.10, 0.29].

In the rewards-known condition, participants estimated the probability to sell a lot of the target fruit at 56.05% ($SD = 23.01$) for the frequent location and at 51.39% ($SD$ = 24.65) for the infrequent location. Again, the estimated did not result in the contingency estimates we had expected. To the contrary, the data supported the hypothesis that participants associated the infrequent location more so with selling a lot of strawberries (grapes) than the frequent location ($\Delta P = 0.05$, $SD = 0.37$), $BF_{-0} = 0.05$, $t(97) = 1.24$, $p = .891$, $d = 0.12$, 95% $CI_d$ [-0.07, 0.32]. See also Figure 4.3.

**Confidence Pre-Measure**

Finally, we asked participants how confident they were in making reasonable estimates regarding the two conditional estimates. They indicated similar confidence in their estimates in the rewards-unknown relative to the rewards-known condition where we would have expected a stronger confidence in the rewards-unknown relative to the rewards-known condition. Specifically, in the rewards-unknown condition they indicated to be 57.32% ($SD = 25.21$) confident in the frequent location and 52.37% ($SD = 25.15$) in the infrequent location. In the rewards-known condition they indicated to be 57.94% ($SD$ = 24.77) confident in the frequent location and 56.73% ($SD = 23.67$) in the infrequent location. The difference in the resulting $\Delta P$s ($\Delta P = 0.05$, SD = 0.271 for the rewards-unknown and $\Delta P = 0.01$, SD = 0.27) for the rewards-known condition did not, however, differ from one another, $BF_{01} = 4.14$, $t(196) = 0.98$, $p = .329$, $d = 0.14$, 95% CI [-0.14, 0.42] .

**Sampling**

On the first trial in which participants were free to choose between both options, 51 out of the 99 participants in the rewards-unknown condition chose the frequent option, $\chi^2(1) = 0.001$, $p = 0.943$. In the rewards-known condition, on the other hand, 61 out of the 98 participants chose the frequent option, $\chi^2(1) = 2.51$, $p = 0.113$.

Across all trials (Figure 4.4), the data, as expected, does not suggest that participants in the rewards-unknown condition chose one option more frequently than the alternative ($M$ = 48.98, $SD$ = 15.08), $BF_{+0}$ = 0.07, $t$(99) = -0.68, $p$ = .749, $d$ = -0.07, 95% $CI_d$ [-0.26, 0.13]. In the rewards-known condition, we expected participants to prefer the frequent location. However, the data did not support this hypothesis ($M$ = 47.84, $SD$ = 13.71), $BF_{-0}$ = 0.68, $t$(97) = -1.56, $p$ = .061, $d$ = -0.16, 95% $CI_d$ [-0.36, 0.04].

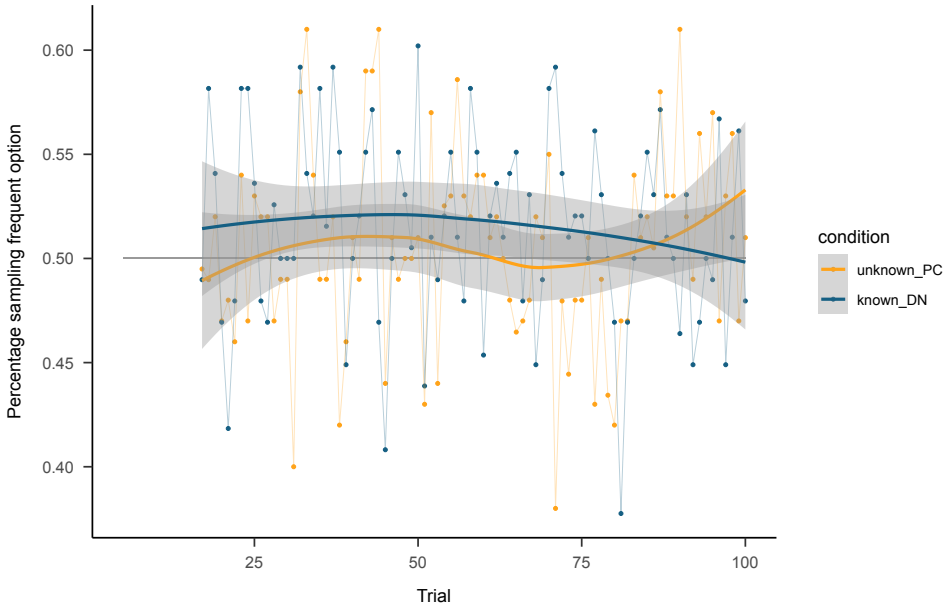



*Figure 4.4.* Percentage of participants sampling the frequent option per trial.

**Relative Contingency Estimate Post-Measure**

After the free sampling phase, we once again asked participants to indicate relative contingency estimates. We had predicted that participants in the rewards-unknown condition would have attenuated initial biases while participants in the rewards-known condition would have maintained their biases. Given that participants in the rewards-unknown condition already reported estimates that did not support the presence of biases on the pre-measure, it will come as no surprise that they once again reported estimates around chance level on the post-measure ($M$ = 0.46, $SD$ = 24.22), $BF_{01}$ = 8.88, $t$(99) = 0.19, $p$ = .85, $d$ = 0.02, 95% $CI_d$ [-0.18, 0.22]. Surprisingly, participants in the rewards-known condition also reported estimates around chance level ($M$ = 0.90, $SD$ = 22.93), $BF_{-0}$ = 0.08, $t$(97) = 0.39, $p$ = .65, $d$ = 0.04, 95% $CI_d$ [-0.16, 0.24].

**Conditional Probability Estimates Post-Measure**

Descriptively, we found an interesting pattern for the conditional probability estimates participants reported. After reporting virtually identical estimates on the premeasure, on the post-measure participants in the rewards-unknown condition reported estimates even closer to chance level, while participants in the rewards-known condition reported estimates that dropped below chance level. While the premeasures are surprising, we predicted this pattern (attenuation in the rewards-unknown condition, biased contingency towards the frequent location in the rewards-known condition) for the post-measures. In numbers, participants in the rewards-unknown condition estimated to have won 58.76% (*SD* = 24.07) of the time with the frequent and 57.27% (*SD* = 24.06) of the time with the infrequent location. Participants in the rewards-known condition estimated to have won 57.51% (*SD* = 20.43) of the time with the frequent and 62.31% (*SD* = 18.54) of the time with the infrequent location. This pattern is, however, descriptive only, as further analyses revealed there to be no statistically significant difference from chance level in either condition. Specifically, the estimated contingency between location and outcome in the rewards-unknown condition was $\Delta P = 0.01$ (*SD* = 0.32), which did not differ from chance level, $BF_{01} = 8.13$, $t(99) = 0.47$, $p = .642$, $d = 0.05$, 95% $CI_d$ [-0.15, 0.24]. Similarly, the estimated contingency in the rewards-known condition was $\Delta P = -0.05$ (*SD* = 0.26) with the data not supporting the presence of a bias, $BF_{-0} = 1.13$, $t(97) = -1.86$, $p = .033$, $d = -0.19$, 95% CI [-0.39, 0.01].

**Confidence Post-Measure**

Finally, and as predicted, participants now did not differ in their estimates regarding the frequent and the infrequent. They estimates their confidence in making reasonable estimates regarding the frequent option (60.65%, *SD* = 27.23) just as high as for the infrequent option (63.21%, *SD* = 27.35), $BF_{01} = 6.46$, $t(193.94) = 0.06$, $p = .952$, $d = 0.01$, 95% CI [-0.27, 0.29].

## Discussion

In two previous experiments, we had found a pattern of initial biases being maintained in what we theorized was an interaction between a primacy effect of the initial evidence and the environment decision makers found themselves in. In reward-rich environments, in which many positive outcomes supported a strategy of premature exploitation, we found that this exploitation led to initial biases being maintained even over an extended sampling period. In reward-impoverished environments on the other hand, many negative outcomes did not invite exploitation. Instead, continued exploration of alternatives led decision makers to attenuate their initial biases (Harris et al., 2020). We extended these findings in a second row of experiments that used a socially more meaningful setting. Here, we found

that biases could even be maintained in reward-impoverished settings depending on how exactly beliefs were formed initially. Different biases led to a focus on different information which then in turn either supported a strategy of exploitation or not (Harris, Fiedler, et al., 2022).

Here, we investigated one potential moderator that we had theorized on previously, namely whether the valence of the outcomes was initially known or unknown to participants. It turns out, at least in the experimental task used here, that valence was not a crucial moderator in explaining when participants formed cognitive biases in line with pseudocontingencies and when they did so in line with a denominator neglect.

Instead, we found that the condition in which the rewards were known from the start and in which we expected cognitive illusions in line with the denominator neglect, participants exhibited the exact pattern we would predict when they formed pseudocontingencies. On the other hand, in the condition in which rewards were initially not known and in which we expected cognitive illusions in line with pseudocontingencies, participants did not exhibit any bias at all. As such, we must be careful in forming any conclusions on the potential moderating role of initially knowing the valence of outcomes. Instead, it is perhaps more likely that our experimental setup did not fully capture the processes at hand. It may very well be that initially knowing the valence of outcomes moderates whether decision makers pay attention to all available information or only selective information. This, in turn, logically would influence what inferences are formed and to what extent further exploitation might uphold these inferences or not. However, it may also be that other processes, such as the difficulty of the task, play a much larger role and the moderating role of initially knowing the valence of outcomes is negligible. Our results speak more to the latter alternative, and we therefore end with the tentative conclusion that initially knowing the valence of outcomes is not, in fact, a strong moderator when it comes to what inferences and cognitive illusions people form based on initial evidence.

# CHAPTER 5

# Chapter 5

## Missing Out by Pursuing Rewarding Outcomes: Why Initial Biases Can Lead to Persistent Suboptimal Choices

**This chapter is based on:**

Harris, C., Aarts, H., Fiedler, K., & Custers, R. (2021). Missing Out by Pursuing Rewarding Outcomes: Why Initial Biases Can Lead to Persistent Suboptimal Choices. *Manuscript submitted for publication.*

## Abstract

While reasons why our initial impressions might be wrong are abundant, further interaction (sampling) opportunities often enough allow us to attenuate such initial biases. Sometimes, however, these biases persist despite repeated opportunities to learn, such as in superstitions or stereotypes. In two studies (*N*s = 100) we demonstrate that in a task in which participants could repeatedly choose between two options to gain rewards, erroneous initial impressions about yielded outcomes can lead to persisting biases towards a clearly inferior option. We argue that a premature focus on reward pursuit (exploitation) rather than exploration is the cause of these biases, which persist despite plenty of opportunities and a presumed motivation to overcome them. By focusing on a supposedly best option, participants never give themselves the chance to sufficiently try out alternatives and thereby overcome their initial biases. We conclude that going for the money is not always the best strategy.

Most human behavior is goal-directed: People act in order to obtain outcomes they find rewarding (Custers & Aarts, 2010). This means that their behavior is for a large part based on beliefs about the relation between actions and their rewarding results. We make a joke because we believe it will cheer someone up, invite a friend over because we want to have a good time, or pick a specific restaurant because we want to have a great meal. While those beliefs may initially be based on suggestions by others (Pilditch & Custers, 2018), or other sources of knowledge, over time they increasingly become based on our first-hand experiences. But what if those initial beliefs were not accurate? Here, we argue that such inaccurate beliefs are not always updated as people pursue rewarding outcomes.

Striving for rewarding outcomes always requires a delicate balance: On the one hand, exploiting the presumably best option may allow for maximizing the intended outcomes in the short run. On the other hand, exploring other options allows for finding and learning about potentially better alternatives. Any decision requires a choice between immediate reward pursuit of a supposedly best option or sacrificing immediate rewards in order to find options that will (hopefully) allow for higher returns in the long run (Cohen et al., 2007; Mehlhorn et al., 2015; Mischel, 1974). Whenever a choice needs to be made, the interests of learning about alternatives and maximizing rewards are pitted directly against one another. Here we argue that pursuit of rewarding outcomes can create a “filter bubble” when it comes to the outcomes people experience. That is, by exploitation of initial beliefs about actions and rewarding outcomes people can create persisting suboptimal biases even when there is an objectively better choice alternative.

As long as decision-makers are correct in their estimations of a given choice environment, exploitation is the optimal strategy. When these estimations are incorrect, however, exploitation would lead to repeated suboptimal decisions. Repeated negative outcomes of such suboptimal choices should quickly discourage further exploitation and instead lead to more explorative behavior. After all, who would continue betting on a losing horse? But when outcomes are repeatedly positive, a decision maker might feel confirmed in their current choice strategy and not notice that other choices would lead to even better outcomes. In other words, positive outcomes might seduce a decision maker into exploiting suboptimal choices without sufficiently exploring choice alternatives (Harris, Aarts, et al., 2022a; Harris et al., 2020). It is an inherent feature of active sampling that the decision-making process constrains the information people receive (Denrell, 2005; Denrell & Le Mens, 2012; Fiedler, 2000; Fiedler & Wänke, 2009). Betting on a winning horse can make us blind to even better alternatives.

In a recent line of experiments, Harris et al. (2022; 2020) have demonstrated such persistent biases in reward-rich environments. Participants played two-armed bandit tasks in which they could repeatedly choose between two options that yielded positive or negative points with the total of these points later being converted to a financial reward. Following a bias induction phase, participants' choice behavior showed overall trends of exploitation across trials (Harris et al., 2020) and at the individual trial level (Harris, Fiedler, et al., 2022), which repeatedly led to persisting biases. However, because the winning probabilities for the two choice alternatives were identical in this line of research, and any biases therefore of no actual consequence, it remains unclear whether these biases could lead people to make suboptimal choices that actually harm them.

In the current research, participants engaged in the same two-armed bandit task used by Harris et al. (2020). Instead of two identical options, however, one option was objectively better than the alternative. In the induction phase, like Harris et al. (2020), we aimed to induce a bias using double-skewed distributions that are known to induce pseudocontingency illusions (Fiedler et al., 2009). More specifically, by presenting one choice option more frequently to participants than the other, and by ensuring that rewarding outcomes for both options occur more frequently than losses, participants should perceive a contingency between the frequent option and the frequent outcome. As a result, people should show an initial preference for the frequent option, even though the true underlying probabilities of obtaining rewards were the same for both options. In the free sampling phase that followed, frequent positive outcomes would then seemingly confirm this initial bias and result in the maintenance of this initial bias throughout an extended sampling period (Harris et al., 2020).

What, then, constitutes a bias in this experimental setting? It is reasonable to assume that participants would not maximize by focusing fully on one (supposedly) better option even though this strategy would lead to the highest probability of positive outcomes. Instead, a common finding is that people match how often they choose options to the probabilities of the outcomes (probability matching; Vulkan, 2000). Probability matching, however, is typically not defined for choice alternatives that are independent of one another. But as probability matching is assumed to emerge from a strategy of win-stay-lose-shift (Nowak & Sigmund, 1993; Otto et al., 2011), we test against the proportions of choices such a strategy would predict. As an even more conservative second test, we also compare participants' behavior to chance level, but note that an indifference between both options would still suggest a bias in that participants clearly haven't picked up on the objectively better option.

## Experiment 1

In this first experiment, we used a distribution of evidence in which the two choice alternatives are still quite similar. While one option resulted in a positive outcome on 75% of the trials, the alternative did so on 80% of the trials.

### Methods

An a priori power analysis for a difference from constant *t*-test using G*Power (Faul et al., 2007) based on measures in Experiment 2 of Harris et al. (2020) suggested a minimum sample size of at least 50 participants for an effect of choice in a condition with frequent positive outcomes for two equal options. These calculations were based on a 5% alpha-level, 80% statistical power, and effect sizes between $d = .35$ and $d = .43$ as reported by Harris et al. (2020) for maintained biases in a condition with positive outcomes on 75% of trials.

Participants for this study were recruited via the online crowd-sourcing platform Prolific Academic (https://prolific.co/) and the study was run in English on Soscisurvey (Leiner, 2020). One-hundred participants ($N_{female} = 41$) with an average age of 27 years ($SD = 7.16$) participated for a financial reward of £0.85 plus additional earnings (mean £1.00, max £1.15) 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. Seventy-eight 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 (19-155). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the experiments (Simmons et al., 2012).

### *Procedure*

The experiment consisted of four phases: An induction phase, in which participants were forced to choose particular options so that they ended up with the distribution of initial evidence that should induce a pseudocontingency (Figure 5.1). Then, a first estimate phase followed in which participants indicated their inferences from the induction phase and which also served as a manipulation check. In a free sampling phase, participants could then choose freely between both options on each trial and earn points that would later be converted to a monetary payoff. In a final estimate phase, participants gave estimates regarding the just completed task. In total, participants completed 100 trials. We counterbalanced which bag was presented more frequently, which color was the winning color, and what color participants were asked to give estimates for.


*Figure 5.1.* Outline of procedure with hypothetical data depicting choices and outcomes (blue = wins, red = losses) for the two choice options A and B.

**Induction phase.** In the induction phase, participants were first introduced to the task consisting of two bags from which they could grab either yellow or blue balls with replacement. Then, they were told that the computer had preselected which of the two bags they would be drawing from on the first few trials in order to get familiar with the task. On a given trial, participants would then, for example, only see bag A. After clicking on this bag and a short delay, they would receive feedback in the form of text ("You chose bag A and drew a yellow ball.") as well as an image depicting, in this case, a yellow ball. After a delay of 1 s the feedback would disappear, and the next choice was presented. Throughout the entire experiment the current trial number as well as the total trial number ("Trial: x/100") were presented on the screen. The induction phase consisted of 17 trials in the distribution of initial evidence presented in Table 5.1. Importantly, while one bag was shown more frequently (12 out of 17 trials), the infrequently shown bag had the higher probability of resulting in a win (80%, vs. 75% for the frequently shown bag). Thus, the co-occurrence of high base rates of winning and the frequent bag should induce a pseudocontingency illusion in favor of the frequent bag, even though the actual winning rate was higher for the other, infrequent bag.

**Pre measures.** Following the induction phase, we asked participants to indicate estimates regarding the distribution they had just encountered. Specifically, we asked them to first indicate a relative contingency estimate ("From which of the two bags were you more likely to grab a yellow ball?") on a slider anchored with the two bags displayed as images at the ends. We then asked participants to estimate for each bag the conditional probability of a yellow bag ("How likely was it (in %) that you grabbed a yellow ball if you chose bag A/B?") on a slider anchored at 0% and 100%. And, we asked them to indicate their confidence in both conditional probability estimates ("How confident are you that you can make a reasonable estimate regarding bag A/B?"). This phase forced participants to actively consider the evidence they had just encountered. Additionally, it provided a straightforward manipulation check of the success of the bias induction.

**Free sampling.** Next, we introduced the reward scheme. From this point on, participants earned 10 points for every yellow and lost 10 points for every blue ball

(or vice versa). Participants were reminded that these points would be converted into a monetary reward at the end of the experiment.

In the free sampling phase, participants were free to choose either bag on each of the remaining 83 trials. Our behavioral measure was the number of choices participants made for each of the two bags during this free sampling phase. 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.

**Post measures.** In the final estimate phase, we asked participants to again estimate the relative contingency, conditional probabilities and their confidence.

Table 5.1

*Distribution of initial evidence*

| | Experiment 1 | | | Experiment 2 | | |
|---|---|---|---|---|---|---|
| | Wins | Losses | | Wins | Losses | |
| Frequently shown bag | 9 | 3 | 75% | 8 | 4 | 67% |
| Infrequently shown bag | 4 | 1 | 80% | 4 | 1 | 80% |
| | | | $\Delta P$ = -.05 | | | $\Delta P$ = -.13 |

*Note*: All percentages depict the ratio of wins out of all trials for the respective location. $\Delta P$ is the difference score between the conditional probabilities and describes the contingency between location and outcome (Allan, 1980)

***Data Preparation***

All sliders were recoded such that positive values in the estimates would indicate a preference for the option that was shown more frequently in the induction phase. In other words, positive values would indicate a pseudocontingency bias, negative values would reflect the true underlying probabilities. The relative contingency estimate had a range of [-50; 50]. We calculated $\Delta P$-values (Allan, 1980), that is, difference scores between two conditional probability estimates of yellow balls given bag A versus B, respectively, which resulted in ranges from [-1; 1]. We did the same for the confidence estimates. Finally, the choice index had a range of [-83; 83].

Data preparation and analyses were undertaken using R (R Core Team, 2018). Across all three measures of preference (the choices made during the free sampling phase, the relative contingency estimates, and the conditional probability estimates), we expected a bias towards the frequently presented option in the induction phase and therefore performed one-tailed *t*-tests and the Bayesian equivalent ($BF_{+0}$). Following a successful bias induction, we expected persisting biases during sampling and on the final estimates and again performed the same one-tailed tests. We expect no difference in participants' confidence in the two conditional estimates and performed two-tailed *t*-tests and the Bayesian equivalent ($BF_{01}$). All graphs include confidence intervals around the means, and we report confidence intervals for the effect sizes.

The objectively better strategy would be to choose the infrequent option, while a successful pseudocontingency induction would result in a preference for the frequent option for which the premeasures served as manipulation check. After finding an overall bias towards the initially frequent, but inferior option, we created a dummy variable for participants based on the estimates from both measures. Specifically, we made one group with all participants whose relative contingency and conditional probability estimates were both larger than zero as would be predicted based on a pseudocontingency inference, the bias group ($n_{bias}$ = 52). The other group, the no-bias group, was made up of all participants that indicated no initial bias in either of their estimates or made estimates in the objectively correct direction ($n_{no\text{-}bias}$ = 48).

In the main text, we present the central findings with regard to participants' choice behavior. Further details on the methods as well as analyses of all variables can be found in appendix D. Additionally, 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 also in appendix D. All analyses and data can be found on an Open Science Framework repository (https://osf.io/2z6sk/?view_only=34df353dbcc848a58cf8fcef5ff71357).

**Results**

Across all trials, participants sampled the initially frequent (but objectively worse) option around 50% of trials ($M$ = .51, $SD$ = 0.29). In other words, participants clearly did not maximize by exclusively choosing the objectively better option. While this was to be expected, participants also chose the initially frequent option more often than a win-stay-lose-shift strategy would predict ($c$ = .444), $BF_{+0}$ = 1.85, $t(99)$ = 2.12, $p$ = .018, $d$ = 0.21, 95% CI [0.01, 0.41]. Instead, participants' choices lay almost exactly at chance level ($BF_{01}$ = 8.90, $t(99)$ = 0.17, $p$ = .862, $d$ = 0.02, 95% CI

[-0.18, 0.21]), which considering that the infrequent option is better indicates biased choices.

To further analyze the data and understand the change of the bias over time, we next split up participants into two groups. Based on their initial estimates, we formed a bias and a no-bias group and further analyzed participants' choices by means of logistic mixed-effects models. Figure 5.2 illustrates the percentage of participants choosing the initially frequent bag per trial.

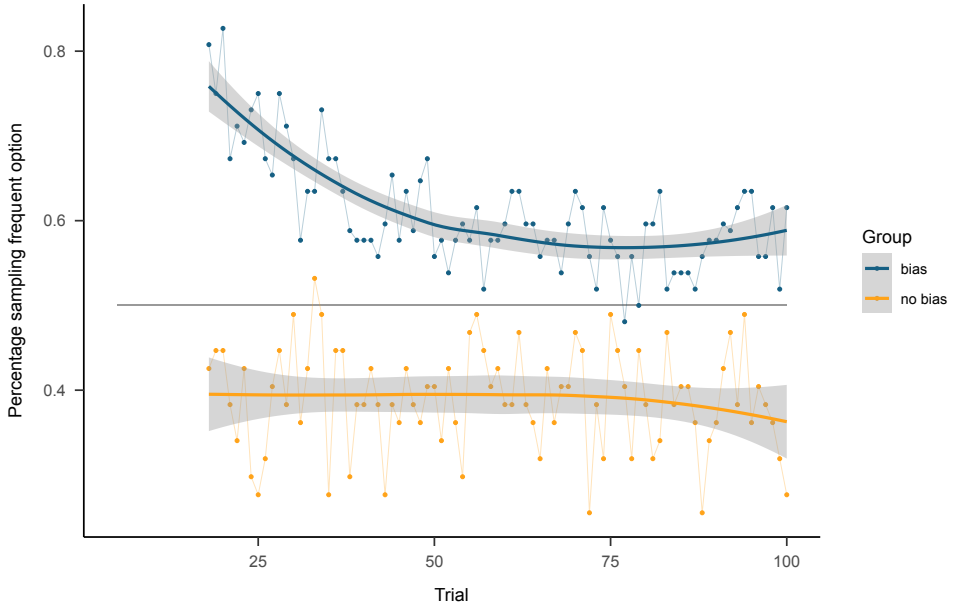



*Figure 5.2.* Percentage of participants sampling the frequent option per trial (Experiment 1).

In the first model, we analyzed participants' choices over time with quadratic general mixed-effects models. In the second model, we included group as a moderating factor. In the third model, we then analyzed the trend at the end of the free sampling period. Due to the binary outcomes, we fitted logistic models to the data. The first free choice trial was set at trial = 0. For model accuracy, we then scaled this number. For the group factor, the no-bias group was coded as 0 and the bias group was coded as 1. Therefore, the first model was:

$$\hat{y} = c + \left(\frac{trial}{100}\right) \times \beta_1 + \left(\frac{trial}{100}\right)^2 \times \beta_2 \tag{5.1}$$

In this model, indicates the log odds[10] for choosing either option. The positive intercept indicated an initial pseudocontingency bias towards the frequent option, $c$ = .61, $z$ = 2.40, $p$ = .017. Across trials, participants increasingly chose the objectively better option as indicated by the negative beta sign, $\beta_1$ = -1.96, $z$ = -4.31, $p$ < .001. However, the positive quadratic effect suggests that this trend mainly occurred in earlier trials, $\beta_2$ = 1.57, $z$ = 3.00, $p$ = .003.

In a second model, we then added the group factor with the no-bias group coded as 0 and the bias group coded as 1:

$$\hat{y} = c + \left(\frac{trial}{100}\right) \times \beta_1 + \left(\frac{trial}{100}\right)^2 \times \beta_2 + group \times \beta_3 + \left(\frac{trial}{100}\right) \times group \times \beta_4 + \left(\frac{trial}{100}\right)^2 \times group \times \beta_5 \quad (5.2)$$

Finally, in order to test for remaining biases on the final trial, we ran the same model but with trial number transformed so that the last trial would be the null point in the model. A negative intercept indicated participants in the no-bias group preferred the infrequent, objectively better option, $c$ = -.89, $z$ = -2.57, $p$ = .010 at the end of the sampling phase. A significant main effect for bias, however, suggests that even after the 100th trial the bias group still differed significantly from the no-bias group, β = 1.84, $z$ = 3.85, $p$ < .001. A fourth model with only the bias group determined that after the 100th trial this group also still differed from chance level, c = 0.96, $z$ = 2.83, $p$ = .005.

Table 5.2

*Regression table for Experiment 1, model 2*

| Effect | Beta | 95% CI | z-value | p-value |
|---|---|---|---|---|
| Intercept | -0.77 | [-1.45; -0.09] | -2.23 | .026 |
| Trial | 0.31 | [-0.93; 1.55] | 0.49 | .621 |
| Trial² | -0.56 | [-2.02; 0.91] | -0.74 | .458 |
| Group | 2.65 | [1.71; 3.59] | 5.53 | < .001 |
| Trial * Group | -4.52 | [-6.29; -2.76] | -5.02 | < .001 |
| Trial² * Group | 4.31 | [2.23; 6.38] | 4.07 | < .001 |

*Note.* Quadratic logistic mixed-effects model, trial number and group as fixed effects, participants as random effects.

10 These can be converted to probabilities using the following formula: $\frac{e^x}{1+e^x}$ where x is the log odd.

**Discussion**

The results demonstrate that a considerable number of participants formed an initial pseudocontingency bias based on the skewness of the initial evidence (Fiedler, 2010). These participants tended to maintain this initial bias throughout an extended sampling phase and confirmed it in a final estimation stage. Despite plenty of opportunities, they seem to have not learned that the alternative option was, in fact, the better of the two options and so their choices led to inferior outcomes. Only participants who had not fallen prey to the skewness of the initial evidence demonstrated a preference towards the objectively better choice option throughout the sampling stage as well as the final estimation stage.

It might seem as if the bias induction was not successful and that any differences between the two groups result from grouping the participants who subsequently failed to learn during the task. However, using the same induction phase, Harris et al. (2020) established a distinct pseudocontingency illusion when the two options were identical. It seems more likely, then, that the objective contingency and the pseudocontingency illusion worked in opposite directions, with some participants still falling prey to the pseudocontingency illusion while others correctly identified the better option as such. Importantly, in what followed participants showed exactly the behavior we predicted: Participants with an initial bias never sufficiently overcame their first impressions. While we observed some adjustments from all participants, these were never sufficient despite more than 4 times the number of trials used to learn the bias initially.



These results demonstrate that initial biases can persist even when the exploited option is actually inferior to alternatives. Reward pursuit can result in behavior that is detrimental to learning. About half of the participants fell prey to the initial evidence and frequent positive outcomes. Ironically, they ended up with less reward than they could have earned exactly because they were so motivated to reap rewards. In the second experiment we increase the difference in expected value between the two choice alternatives further to investigate whether more extreme distributions will allow participants to overcome their initial biases more readily.

## Experiment 2

In the second experiment we use a more extreme distribution of evidence. Now, the initially frequently presented option results in a positive outcome on 67% of the trials whereas the alternative results in a positive outcome on 80% of the trials.

**Methods**

Participants for this study were again recruited via Prolific Academic (https://prolific.co/) and the study was run in English on Soscisurvey (Leiner, 2020). One-hundred participants ($N_{female}$ = 58) with an average age of 31 years ($SD$ = 9.25)

participated for a financial reward of £0.85 plus additional earnings (mean £1.00, max £1.15) 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. Eighty-five percent of participants had an educational degree of College/A levels or higher. The distribution of initial evidence used in the induction phase is presented in Table 5.1.

**Results**

Across all trials, participants sampled the initially frequent (but objectively worse) option around 40% of trials ($M$ = .40, $SD$ = 0.27). Participants still did not maximize by exclusively choosing the objectively better option. Now, however, their choices no longer differed from our win-stay-lose-shift benchmark ($c$ = .377 ;$BF_{01}$ = 5.51, $t$(99) = 1.01, $p$ = .315, $d$ = 0.10, 95% CI [-0.10, 0.30]) and were therefore lower than chance level, $BF_{10}$ = 36.51, $t$(99) = -3.56, $p$ < .001, $d$ = -0.36, 95% CI [-0.56, -0.15].

Splitting participants into groups based on the successful bias induction, however, again reveals a striking pattern that is visualized in Figure 5.3 and which we describe analytically using the same mixed-effects models as in Experiment 1.

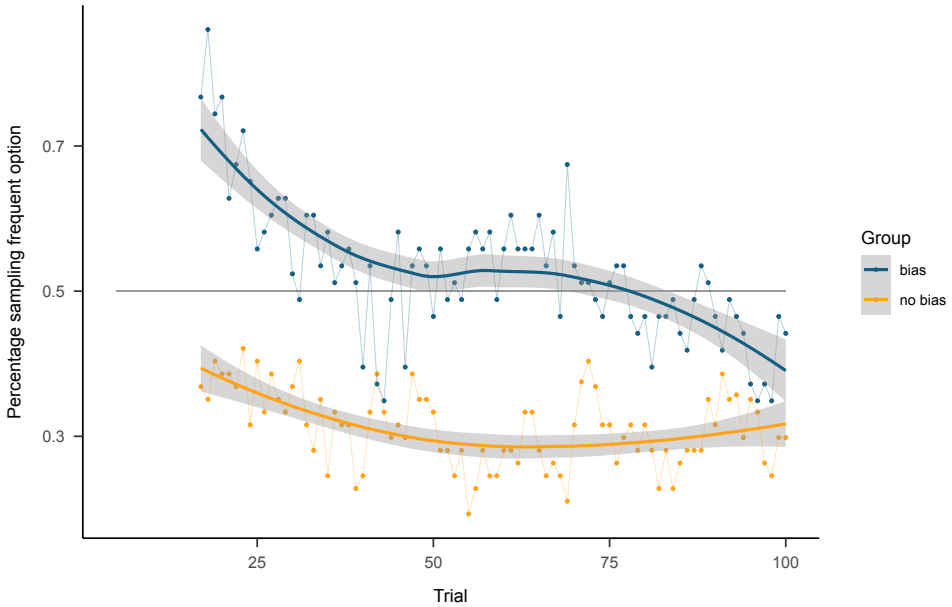



*Figure 5.3.* Percentage of participants sampling the frequent option per trial (Experiment 2).

In the first model, in which we compared participants' choices over time, we did not find any indication for an initial bias across all participants, $c = .02$, $z = 0.09$, $p = .926$. The negative main effect for trial number again indicated a tendency to increasingly often choose the objectively better option ($\beta_1 = -2.55$, $z = -5.90$, $p < .001$); the quadratic main effect suggested this trend to be more extreme in the beginning of the sampling phase, $\beta_2 = 1.96$, $z = 3.89$, $p < .001$.

In the second model, we then again included group as a factor with the no-bias group as reference. The negative intercept indicated a preference of the reference group, that is, the no-bias group, to choose the initially infrequent, objectively better option, $c = -0.78$, $z = -2.78$, $p = .005$. This intercept matches almost exactly the win-stay-lose-shift benchmark of $c = .333$ (). The negative main effect for trial number indicated that participants in the no-bias group increasingly chose the initially infrequently, objectively better bag, $\beta_1 = -2.48$, $z = -4.24$, $p < .001$. The positive quadratic main effect indicated that this negative effect was strongest in the beginning of the sampling phase, $\beta_2 = 2.42$, z = 3.53, $p < .001$. The positive main effect for group reflects the biased group's preference for the frequent option, $\beta_3 = 1.86$, $z = 4.36$, $p < .001$. We report the full model in Table 5.3.

To test for choice trends on the latest trials, we ran the same model with trial number transformed so that the last trial would be the null point in the model. As in Experiment 1, a negative intercept indicated participants in the no-bias group preferred the infrequent, objectively better option at the end of the sampling phase, $c = -1.36$, $z = -5.08$, $p < .001$. A significant main effect for bias, in contrast, reflects a significant difference between the bias group and the no-bias group even after the 100th trial, $\beta = 1.06$, $z = 2.50$, $p = .012$. A fourth model with only the bias group determined that after the 100th trial this group no longer differed from chance level, c = -0.12, $z = -0.41$, $p = .679$.

Table 5.3

*Regression table for Experiment 2, model 2*

| Effect | Beta | 95% CI | z-value | p-value |
|---|---|---|---|---|
| Intercept | -0.78 | [-1.33; -0.23] | -2.78 | .005 |
| Trial | -2.48 | [-3.63; -1.34] | -4.24 | < .001 |
| Trial$^2$ | 2.42 | [1.07; 3.76] | 3.53 | < .001 |
| Group | 1.86 | [1.02; 2.70] | 4.36 | < .001 |
| Trial * Group | -0.22 | [-1.92; 1.49] | -0.25 | .804 |
| Trial$^2$ * Group | -0.89 | [-2.88; 1.09] | -0.88 | .379 |

*Note.* Quadratic logistic mixed-effects model, trial number and condition as fixed effects, participants as random effects.

**Discussion**

In this second experiment we used distributions of initial evidence that differed more extremely in their expected value than in Experiment 1. Even then, we found that initial biases persisted extraordinarily long and were only attenuated around the 75th trial. But even after the 100th trial, the estimates participants made (see appendix D) indicate that there still remained doubt as to whether the objectively better choice alternative was indeed better. The results once again demonstrate that it was not participants' unwillingness to learn that results in the maintenance of biases. Rather, when participants exploit one option, any alternative has to be markedly better for participants to notice.

As in any situation with two choice options in which one has better odds, maximizing the better option results in the highest payoffs (Hinson & Staddon, 1983). Even participants without an initial bias as a group did not follow this strategy, instead showing behavior reminiscent of probability matching (Vulkan, 2000)[11]. But participants with an initial bias did not even follow this less optimal but often adaptive strategy (cf. Gaissmaier & Schooler, 2008). Instead, the cognitive illusion overrode the genuine contingency across both experiments and participants ended up forfeiting the higher reward probabilities the objectively better option had to offer.

**General Discussion**

In two experiments, participants tried to earn as many points as possible to gain financial rewards by repeatedly choosing between two options. In an induction phase, one option was presented more frequently while the other, infrequent option resulted in a higher winning probability and was therefore the objectively better choice. About half the participants mistook the frequency of presentation to imply a contingency and incorrectly inferred that the more frequently presented option was better (Fiedler et al., 2009). These participants maintained this bias throughout the extended free choice phase and even on later estimations thereby forfeiting better winning probabilities and ultimately financial rewards. Only participants who did not initially fall prey to this cognitive illusion correctly preferred the objectively better choice option later on.

We believe that an initial bias tempts premature exploitation and frequent wins seemingly confirm the current strategy. Why change a winning horse, as the saying goes? But by exploiting a seemingly best option, decision makers deprive themselves of opportunities to learn about choice alternatives. Intriguingly then, they end up with very idiosyncratic evidence in which one option was chosen most of

11 Note though that probability matching is not clearly defined in this case as both choice options are independent of one another.

the time and alternatives are never given a fair chance – people's choices lead to a highly subjective representation of the world.

Our findings demonstrate the dramatic effects this initial experience can have on later decisions. Of course, other processes (in particular motivational; Kunda, 1990) influence one's information search and belief-updating as well and perhaps even more extremely, which poses limitations on the generalizability of the findings and exciting avenues for further research. Nonetheless, the current paper is in line with a growing literature that has put increasing emphasis on belief-updating in the light of continued evidence (Alves et al., 2018; Bott & Meiser, 2020; Harris et al., 2020; Pilditch & Custers, 2018) and on the role of active sampling in decision making (Denrell & Le Mens, 2012; Fiedler & Wänke, 2009; Prager et al., 2018). Decisions are rarely made in a vacuum, but instead reflect the history of the decision maker. But this history is highly subjective. It depends on earlier actions which in turn depend on beliefs. Yet, each decision, in turn, limits to what extent new information can be learned and beliefs can be updated (Denrell, 2005; Harris, Aarts, et al., 2022a).



This notion of the subjective experience can explain a wide range of social phenomena. Why do some people believe so strongly in the effectiveness of certain alternative medicines while the medical sciences at best find mixed evidence? One's subjective experience might suggest a strong contingency: When I have the flue, I take my remedy, and within a few days I feel better. People want to get healthy quickly and so they rely on what seems to have worked in the past. Few people, when sick, are willing to engage in careful hypothesis testing. But not only are some people using less effective treatments, the costs of such behavior might be considerable for health care systems.

Perhaps even more striking are the effects such persisting biases can have in social interactions. Initial biases about, for example, the trustworthiness of interaction partners will lead to either engagement or to their rejection and affect to what extent biases can be updated or remain unchanged with detrimental consequences for interaction partners. By being shunned they never receive the opportunity to disprove the existing stereotypes and remain systematically disadvantaged (Jaeger et al., 2021). The current research demonstrates that biases may persist even when they are of disadvantage to the individuals holding the biases.

Finally, these processes might even be at the heart of stereotype maintenance. Existing stereotypes will factor in when deciding whom to interact with in social settings. After all, when seeking rewarding interactions, one would rely on their beliefs about and impressions of others. And as long as one encounters sufficient rewarding interactions in the majority group there might not be any reason to doubt one's strategy. As a consequence, however, instead of reducing

the discrepancy in information and interaction with minority group members that often leads to stereotypes in the first place, this imbalance is maintained or even strengthened (Alves et al., 2018; Denrell & Le Mens, 2011; Kutzner & Fiedler, 2017). It comes as no surprise then, that interventions often focus on increasing contact between majority and minority groups (Pettigrew & Tropp, 2006). But how can stereotypes persist also in situations in which there are opportunities for contact between majority and minority groups? The current research suggest that we might not be taking advantage of such opportunities because we focus too strongly on the supposedly best choice alternatives, the majority group we know well, and neglect alternatives, such as less known minority groups.

In conclusion, one's information-sampling strategies induced by one's early experience can markedly constrain the extent to which beliefs can ever be updated. This can lead to discrepancies between one's subjective experience and the objective world, the consequences of which can lead to beliefs and choices that are harmful to the individual, interaction partners, and society as a whole.

# CHAPTER 6

# Chapter 6

## Iterative Cycles in Psychology: How Hedonic Outcomes, Idiosyncratic Experiences, and Decisions by Experience Shape Our Subjective World

**This chapter is based on:**

Harris, C., Aarts, H., Fiedler, K., & Custers, R. (2021). Iterative cycles in Psychology: How Hedonic Outcomes, Idiosyncratic Experiences, and Decisions by Experience Shape Our Subjective World. *Manuscript in preparation.*

## Abstract

In order to realize desirable and rewarding outcomes in changing environments, people have to constantly decide what choice alternatives to engage with. This process involves three stages: encountering evidence, updating one's beliefs, and making the next decision. Here we argue that these stages are highly dependent on one another and together form an iterative cycle. Within this cycle, people's choices necessarily lead to an idiosyncratic experience of the world that ideally offers a (fairly) accurate representation of the environment and desirable action-outcomes therewithin, but that can easily become biased. Such biased representations can explain persisting erroneous beliefs and suboptimal goal-directed behaviors. In the present paper, we discuss recent evidence for such biases as a result of this iterative cycle and discuss possible opportunities for interventions that help people to create a more objective representation of the outcomes of their actions.

The navigation of our environments depends on a myriad of decisions that need to be made: Whom do we trust to be cooperative? Whom do we seek out for social connections and positive experiences? And where should we go for the best coffee in town? The more we zoom in, the more apparent it becomes just how many decisions we constantly have to make in order to achieve our goals and to attain desirable outcomes. It should therefore come as no surprise that humans, by and large, are adept decision makers. We tend to be quite sensitive to covariation or patterns in our environment and canonical literatures on learning and belief-updating speak to our capabilities to learn from our experiences, update our beliefs, and adjust future decision-making (Fiedler & Wänke, 2009; Griffiths & Tenenbaum, 2006).

Nonetheless, there are also many examples in which exactly this learning and updating of beliefs seems to fail: People hold prejudices about other people or groups of people, develop superstitious rituals, and have countless other inaccurate or false beliefs about their environment. In many Western countries, for example, many people believe in treatments using certain alternative medicines, despite difficulties in the medical sciences to find convincing evidence for their effectiveness (Relton et al., 2017; Shang et al., 2005). Interestingly, in all of these examples, individuals typically have plenty of opportunities to put their beliefs to the test, if not immediately then at least over time. And in the medical example one could certainly assume that they would also be strongly interested in finding the best, that is, most effective, treatment. And yet, we see a huge divide between people's beliefs and the evidence found in rigorous scientific investigations. Why, then, do such false or inaccurate beliefs persist in the course of goal achievement?

Obviously, learning and belief-updating rely on aggregating (good) information and as such are strongly dependent on the distribution of information in the environment. While beliefs are often formed and updated based on information offered by others (e.g., the media, education, peers, etc.), they are also formed and updated based on first-hand experiences. More specifically, humans are not only passive consumers of information, they also actively engage with their environment, pursuing positive experiences and learning along the way. Well-established literatures detail how encountered evidence can lead to belief-updating (e.g., what consequences our actions will have; Griffiths & Tenenbaum, 2006)what consequences our actions will have; Griffiths & Tenenbaum, 2006. And similarly, well-established literatures detail how our beliefs then, in turn, affect our reason-based behavior and the choices we make (Ajzen, 1985; Gollwitzer, 1990). Here, we highlight a third stage in the process: how our behavior affects what evidence we can possibly encounter. When all three stages are considered, we end up with an iterative cycle of decision-making that stresses the interplay between the stages

and how they may lead to conclusions that reflect the idiosyncratic evidence one encountered but that are an incorrect reflection of the true underlying contingencies of the environment. Such a framework can explain the persistence of biases and it describes how our choices ultimately shape the world we perceive. At the end of the chapter, we will demonstrate how this perspective may help us to understand and explain persistent biased that are observed in various domains of psychology.

In the next parts, we first outline why each stage is important to consider and how it affects the next stage. Then, we describe specifically a process which results in our iterative cycle that is based on sampling in pursuit of hedonic outcomes. We describe this cycle against the backdrop of social cognition literature. In the end we argue that this cycle can be generalized to other types of sampling such as the avoidance of negative outcomes.

**Inferences, Choices, and Sampling**

Learning plays a vital role in human life. It is omnipresent. And yet, learning in and of itself is rarely what humans strive for. Instead, learning is usually a functional aspect in goal pursuit and the maximizing of rewards. For instance, many people in the workforce might strive for an efficient commute to their work. They might deviate from their typical route for a number of reasons (e.g., because of construction along the way or because one believes a different route or means of transportation might be faster), but it happens less often that people explore alternatives for the sake of it. Learning, then, can be seen as a by-product of goal-pursuit, based on the feedback one receives as the result of their actions. And while humans can learn from observing others (Bandura & Kupers, 1964) and even by communicating beliefs and instructions (Pilditch & Custers, 2018), the critical test remains the first-hand experience they gain; that is, the outcomes they experience as they engage with the environment, sampling information from it in order to obtain desirable outcomes. Because this form of aggregating knowledge depends on the idiosyncratic experiences one makes, it is sometimes referred to as hedonic sampling or experience sampling (Denrell, 2005; Denrell & March, 2001; Eiser et al., 2003). While important distinctions can be made what a desirable experience specifically entails (e.g., Berridge, 2000; Berridge & Robinson, 2016), we use a very broad definition here. We will use rewards as a simple term to describe all experiences from immediately pleasurable experiences to long-term pursuits of desirable outcomes that the decision-maker tries to obtain.

***The Exploration-Exploitation Tradeoff***

Decision-makers continuously face a fundamental tradeoff in hedonic sampling: Do they still have to learn more information about their environment in order to find (the most) rewarding outcomes? Or have they already aggregated sufficient information to maximize positive outcomes from a presumably best option?

In other words, should they follow a strategy of exploration or exploitation (Cohen et al., 2007; Hills et al., 2015; Mehlhorn et al., 2015; Wilson et al., 2014)? Exploration is characterized by trying out choice alternatives and is, of course, essential to orientate in a new environment or to pick up on changes in an environment such as changes in outcome probabilities. As a logical consequence, however, exploration can result in making suboptimal choices and suboptimal long-run outcomes. Exploitation, on the other hand, is characterized by maximizing positive outcomes by engaging with a presumably best option. As a logical consequence, however, exploitation can result in insufficient sensitivity to changes (e.g., in outcome probability) that choice alternatives undergo.

Whenever decision-makers face repeated decisions with partial feedback (Hertwig & Erev, 2009), they face this tradeoff and have to balance the potential benefits of further exploration with the potential benefits of exploitation. Naturally, this tradeoff is not an all or nothing distinction: decision-makers are known to often balance exploration and exploitation, for example by matching their choice probabilities to the outcome probabilities of interest (probability matching; Vulkan, 2000). Nonetheless, even with more balanced strategies, exploitation inherently implies a stronger focus on certain options over others.



***First Impressions***

Obviously, there are situations that call for exploration such as an unknown environment, a particularly dynamic environment with many changes (Biele et al., 2009), social cues (Winet et al., 2020), or also when the decision-maker has a clear epistemic goal. But we argue that more often than not, decision-makers are mainly interested in reward pursuit. Exploitation then can be an extremely successful strategy – as long as one's representation of the choice environment is sufficiently accurate. This, however, is not a trivial assumption.

Particularly first impressions are incredibly difficult to get right and the many ways in which decision makers can err are well-documented throughout the literature: They are fallible for order effects and, for example, tend to overweight initial evidence over later evidence (primacy effects; Anderson, 1965; Asch, 1946; Dennis & Ahn, 2001; Jones et al., 1972). They tend to take probabilities and encountered evidence at face value while neglecting the underlying base rates (cognitive myopia; Fiedler, 2012). And, they tend to use cognitive algorithms for inferring contingencies that are not always logically warranted (e.g., pseudocontingencies; Fiedler et al., 2009; Fiedler et al., 2013). But even without particular cognitive biases, people's initial inferences may be biased due to the skewness of small samples (Hertwig & Pleskac, 2010; Kareev, 2000; Kareev et al., 2014) or even simply due to random fluctuation in

the environment. Again, first impressions that are formed in order to understand and utilize the environment are actually extremely difficult to get right.

***Choice Behavior***

These first impressions, however, then serve as guide for later behavior. As agentic beings, people oftentimes engage actively with their surroundings in order to satisfy needs and obtain goals (Custers & Aarts, 2010). It goes without saying then, that preferences and judgements are the compasses people steer by when they choose or act. Expectancy-Value theories stress that people would engage in the behavior that has the largest expected value based on previous learning (Edwards, 1954; Kahneman & Tversky, 1979; Savage, 1954; Tolman, 1934; Von Neumann & Morgenstern, 1947). Other theories, such as the Theory of Reasoned Action (Fishbein & Ajzen, 1975) and the Theory of Planned Behavior (Ajzen, 1991), consider attitudes (based on previous behavior) as one of the main influences on intentions and behavior. As such, the first impressions (but also, of course, later impressions) people learn from their experiences extend a direct influence on behavior. And they do so whether one's beliefs are biased or not.

**Biases in Experience Sampling**

***Hedonic Sampling***

The critical consequence, however, is that as people engage in certain behavior based on their (biased) initial beliefs, feedback will oftentimes be contingent on the behavior they exhibit. In other words, by engaging in some choice alternatives, they can only learn about these options but not others. As a consequence, they end up with a subjective subset of information that can differ from the ecology's true probabilities (Dawes et al., 1989; Denrell, 2005; Denrell & Le Mens, 2011). This is perfectly illustrated by an example introduced by Denrell (2005): Assuming that a strong criterion in choosing whom to interact with is the experience of positive interactions (cf. Thorndike, 1927), positive first impressions increase the chance of further interactions. This in turn would allow one to learn more about the other person and thus update the first impression to a more accurate estimate – be it positive or negative. If the first impression is negative, however, the chance of further interactions decreases. The logical consequence then is that these negative first impressions are less likely to get updated, regardless of whether they were accurate or not. Seeking out hedonically positive interactions with others thus leads to an overall negativity bias because false negative first impressions are less likely to be followed up on (and updated) while false positive impressions do tend to get corrected.

Further research by Denrell and Le Mens (2011) specifies under which conditions such illusory correlations will be positive or negative. This example

highlights how one's beliefs and resulting behavior directly affect the feedback one receives from the environment and how they can limit to what extent belief-updating is even possible. As long as people's inferences are reasonably accurate, it makes perfect sense to base one's decisions on one's beliefs. But when (initial) inferences are biased, the common assumption that people will attenuate initial biases as they continue sampling might not be so straightforward.

***Confirmation Bias and Positive Testing***

This interplay between one's beliefs and decisions may be familiar from the confirmation bias literature (Klayman, 1995). A seminal paradigm in this field was introduced by Lord et al. (1979) in which participants read vignettes confirming or disconfirming their beliefs regarding the death penalty. Participants rated vignettes that confirmed their views on this topic rather than opposed them as more convincing. In other words, they would weight incoming information differently in order to uphold worldviews they held, to protect their self-image, or for other motivational reasons (Klein & Kunda, 1992; Kunda, 1990). Here, an ulterior goal can lead to the biased integration of potentially unbiased information which in turn affects belief updating.



However, when setting out to test a given hypothesis, people are also known to search for confirming evidence more so than for disconfirming evidence (cf. positive testing; Fiedler et al., 1999; Klayman & Ha, 1987). A particularly vivid example of positive testing is the practice of ethnic profiling in which police use ethnic characteristics as cues in their decisions, for example whom to stop. By basing this decision on ethnic characteristics such as the complexion of people's skin color, one would inevitably aggregate skewed evidence: If more people of a particular complexion are searched, a higher absolute number of hits will be found even when assuming equal base rates. In other words, the directed information search will tend to result in a disproportionate absolute amount of confirming evidence. And this in turn affects how beliefs are updated following this information search. That is, behavior is contingent on beliefs, and feedback in turn is contingent on behavior. But the updating of beliefs is obviously dependent on encountered feedback and as such biased initial beliefs might perpetuate in an iterative cycle leading to maintenance rather than attenuation of the initial belief as people continue sampling. Motivational accounts of confirmation bias that stress biased information integration as outlined above would only further accentuate this process.

While the confirmation bias and positive testing can both lead to persisting biases because beliefs influence behavior, there is an important distinction to be made to hedonic sampling. Confirmation bias and positive testing are both processes specifically related to the goal of trying to test hypotheses while hedonic sampling describes the process of making choices in pursuit of positive outcomes such as

rewards. As a result, confirmation bias and positive testing could both involve seeking out negative options if one were to test the hypothesis that a given option is indeed negative. Hedonic sampling and more specifically reward pursuit, on the other hand, would entail making choices that (short- or long-term) result in positive outcomes.

**The Iterative Cycle of Decision-Making**

The central theme here is the iterative nature of experience sampling: Beliefs influence behavior which in turn influences feedback and feedback in turn influences beliefs; where beliefs refer to the mental representation of (expected) covariance in the environment, behavior refers to the decisions we make to engage with some options over alternatives, and feedback refers to the information or outcome that results as a consequence of the behavior displayed. The interplay of these three stages may lead to conclusions that correctly reflect the idiosyncratic evidence encountered, but that incorrectly reflect the true underlying contingencies of the environment (cf. Denrell, 2005; Denrell & Le Mens, 2011; Fiedler, 2000; Fiedler & Wänke, 2009). As such, a comprehensive framework of decision-making would need to account for the active role a decision-maker has in sampling and the information learned, which is contingent on previous decisions that were made. One's own, subjective experience that is built during repeated interactions would need to play a central role in such a framework. Figure 6.1 illustrates this iterative cycle of learning during reward-pursuit.

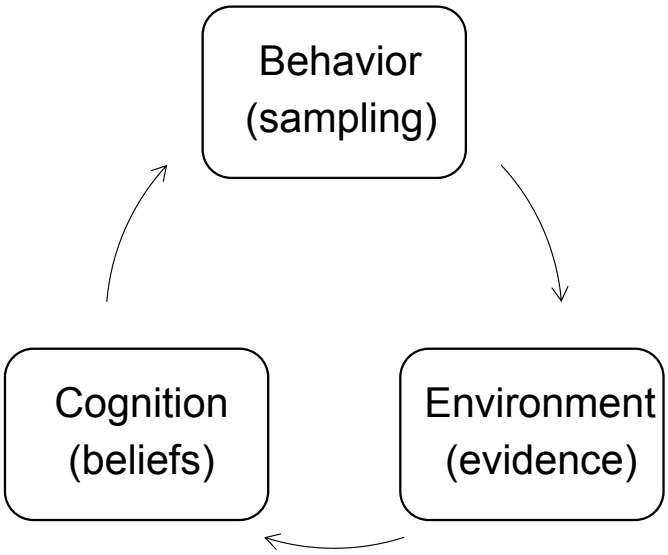



*Figure 6.1*. The iterative cycle of decision-making

***Outside Influences***

Each of these three stages is susceptible to outside influences that not only can lead to biases at this particular stage but that can lead to perpetuating biases. Ask yourself what your favorite pub in town is and how you came by that conclusion. While there are plenty of good reasons why one might prefer one pub over others, here we want to discuss three reasons that relate directly to the iterative cycle we just introduced, namely guided sampling through other agents, particularly good or bad experiences, and external recommendations.

Imagine, for example that a friend you visit always takes you to Irish bars. Your choice behavior, what bar you go to, is determined externally through your friend's preference (Le Mens & Denrell, 2011). But by frequenting some type of pubs and avoiding others, the information you learn about the pubs in town will obviously be skewed. This, in turn, will affect to what extent you can possibly update your beliefs – which will have consequences for future choices. Your behavior can lead to persisting biases.

Alternatively, imagine you go to a new bar and, because of the crowd that night, you have a particularly good or bad night out. Analogous to the earlier mentioned example by Denrell (2005), this will likely affect your future choices and in particular whether you will return to this bar. A bad experience will likely decrease the chance that you will return in the future, while a particularly positive experience might increase this chance drastically and overshadow alternatives. The valence in outcomes at the evidence stage obviously affects belief-updating. Feedback, that is, the evidence you encounter, can again lead to persisting biases.

And finally, imagine that you find yourself in a new part of town and have received ideas from friends or off the internet about good establishments in the area. A belief that one pub is particularly promising (or to be avoided) will affect your decision to go to some pub and make experiences there but not at other pubs. Here, the belief stage affects choices which will determine the information that can possibly be learned. Beliefs, too, can lead to persisting biases.

A preference for certain types of pubs, a good (or bad) night out, or recommendations by others – all of these can affect not only single decision-making instances but can lead to long-term (biased) preferences. At the trial level, the choices made in each of these examples seems perfectly reasonable: One is out for a good night, why wouldn't one base this decision on preferences, past experiences, or recommendations? But at the aggregate level, that is, across multiple such decisions and when the decision-making and belief-updating processes repeatedly run through the iterative cycle, the resulting evidence can easily become skewed such that mental representations that accurate reflect the environment can become

impossible to obtain. In conclusion, outside influences can affect one's beliefs, behavior, or the evidence encountered in ways that can lead to persisting biases.

***Reward Pursuit and the Self-Sustaining Cycle***

Our tendency as humans to seek out positive and rewarding situations (Thorndike, 1927), means that there is an increased chance that certain positively-associated options are chosen over other negatively-associated options. The motivation to seek out positive experiences when entering the choice stage has the effect that feedback one receives will be skewed such that the sought-out choice alternatives are overrepresented relative to alternatives that are avoided. And, as argued above, this restricts to what extent one's beliefs can possibly be updated – with down-stream consequences for future decisions.

More often than not, reward-pursuit, or more generally a strategy that favors exploitation over exploration, will lead to a sample of evidence that is skewed. While such a strategy is likely to lead to higher chances for positive outcomes, it can make optimal learning difficult. The sampling stage, evidence stage, and belief stage are highly interdependent. And as a consequence, reward-pursuit can result in illusory correlations and an idiosyncratic experience of the environment that is maintained during continued interaction (Denrell, 2005; Denrell & Le Mens, 2011; Denrell & March, 2001). Recent experiments demonstrated empirically that as long as they encountered a sufficient number of positive outcomes overall, participants maintained initial biases between two or more equally good choice alternatives (Bai et al., 2021; Harris, Fiedler, et al., 2022; Harris et al., 2020; Kasper et al., 2021) and even when the bias was towards the worse of two choice alternatives (Harris, Aarts, et al., 2022b).

The pursuit of positive outcomes results in focusing on some supposedly best options. Despite plenty of opportunities to put one's beliefs to the test, (premature) exploitation and reward pursuit imply that these opportunities are not made use of as instead the focus is on maximizing positive outcomes. The result is an idiosyncratic experience and within this idiosyncratic experience a maximizing strategy might seem perfectly reasonable. After all, as long as one believes a certain option to be best, surely this option should be chosen? But at the aggregate level, the resulting evidence will be skewed which can make it extremely difficult to notice that the currently preferred option is not, in fact, the best available option. In other words, frequent choices for a supposedly best option will result in an overall distribution in which this option is overrepresented. This, in turn, can make it increasingly difficult to detect that one's preferred choice option is not, in fact, the best option available (cf. Harris et al., 2020). A decision-maker could be pursuing other strategies (e.g. satisficing; Simon, 1955, 1956) thereby allowing for more freedom in choices and

mitigating the effects of a maximizing strategy. But the important notion here is that the strategic orientation towards reward pursuit seldom aligns with behavior that would result in optimal learning and vice versa. Hedonic sampling rarely goes hand in hand with the type of sampling necessary for optimal learning. Ironically, then, people forgo better alternatives exactly because they are so focused on maximizing their rewards.

***Self-Sustaining Cycles in Social Cognition***

Iterative cycles based on hedonic sampling underlie several phenomena in social cognition, such as illusory correlations, the confirmation bias and positive testing, ingroup/outgroup effects, and the halo effect among others (for discussions see Denrell, 2005; Denrell & Le Mens, 2011; Fiedler, 2000; Fiedler & Wänke, 2009). One particularly illustrative example concerns the formation and maintenance of erroneous beliefs and stereotypes that we have of others when deciding whether to interact with them or not, for example when cooperation could lead to higher rewards or interaction leads to a pleasant experience.



As already argued above, an important source of information that can result in erroneous beliefs of others are the first impressions we form. For these, among many other cues, we tend to rely on physical, and in particular facial features, far beyond their cue validity (Aviezer et al., 2015; Feingold, 1992; Wilson & Eckel, 2016). Nonetheless, these (facial) features influence our first impressions and consequentially also our future decisions. As discussed above, this can lead to persisting biases. In a particularly straightforward empirical demonstration, Jaeger et al. (2021) had participants play multiple rounds of a trust game with interaction partners (computer-generated faces that varied in how trustworthy they looked and in how trustworthy they truly were). Participants sought out trustworthy-looking partners and, by interacting with them, learned whether these partners were truly trustworthy or not. Untrustworthy-looking partners, on the other hand, were shunned and by never interacting with them, participants never gave themselves the opportunity to learn whether their initial assessment was accurate or not (see also Fetchenhauer & Dunning, 2010).

This is as an extreme case of the cycle we are advocating: No contact at all means no further information and therefore no belief-updating. But it is easy to see how also more nuanced behavior tendencies can lead to persisting biases (e.g., based on the first few trials; Pilditch & Custers, 2018; Staudinger & Büchel, 2013) – and also how other social cues can induce and maintain this cycle (e.g., race; Burns, 2006). In fact, this mechanism, in which people shape the evidence they perceive by their goal-pursuit, also plays a central role in the perpetuation of one-sided or even biased information distributions – in particular on social media. After watching a few

cat videos or fail compilations on an online video platform, the algorithm will start suggesting more and more such videos. In fact, it too capitalizes on exploitation: The algorithm's goal is to keep people engaged for as long as possible and so it will present content that will keep the viewer hooked – thereby maximizing the viewer's time spent on the platform. Similarly, one's interests on Twitter, (e.g., funny content vs. serious discussions of work-related topics) will likely lead them to follow certain people more than others and the platform to suggesting certain content more often to them than other content. Hence, individual preferences create a completely different and highly idiosyncratic platform experience (Geschke et al., 2019).

Both online and offline, our choices determine what feedback we might possibly encounter and, critically, these interactions restrict to what extent we can update our beliefs. It should come as no surprise then, that the most canonical intervention to stereotype beliefs is based on creating contact and thus opportunities for learning new information about others and updating one's beliefs (contact hypothesis; Pettigrew & Tropp, 2006) and that both Denrell (2005) and Jaeger et al. (2021) discuss contact, proximity (Ebbesen et al., 1976; Preciado et al., 2012), or, more generally, sampling opportunities as potential intervention to these persisting biases. The contact hypothesis has led to a substantial body of research on intergroup contact and has repeatedly led to policy recommendations (Paluck et al., 2018; Paolini et al., 2021; Pettigrew & Tropp, 2006). And indeed, creating contact opportunities can help alleviate erroneous impressions and stereotype beliefs (Pettigrew & Tropp, 2006, 2008).

However, whereas this intervention is seemingly effective, it is crucial to highlight an important distinction that results logically from our comprehensive cycle. Namely, that creating these opportunities alone does not solve the issue – it is equally important that decision-makers are inclined to actually engage with these opportunities (Kauff et al., 2020) by means of their behavior (i.e., their choices) and their willingness to update their beliefs (as opposed to e.g. confirmation bias). Harris et al. (2020) demonstrated that initial biases can persist even when opportunities for putting one's beliefs to the test are available. When decision-makers do not engage with these opportunities (because they prefer a trustworthy-looking interaction partner in a trust game, want to hire the supposedly best candidate, or simply because they want to earn as much money as possible in an experiment), the opportunities in and of themselves are worth little. It makes sense that policies aimed at creating opportunities for contact could lead to positive intergroup contact. But it is also quite possible that each group sticks to their own because this is where they expect the most positive experiences (see also Kauff et al., 2020).

Such considerations might be difficult to spot if one only investigates an isolated stage in the decision-making and belief-updating cycle. But a comprehensive view as we propagate it here makes these interdependencies apparent: It is not only about a decision-maker's motivation to engage with certain choice alternatives, and it is not only about available opportunities. It is about the interplay of all three stages, that is, one's beliefs, one's behavior, and the evidence one encounters. In the next section we discuss consequences for interventions that result from our comprehensive framework.

**Conclusion and Implications**

In the present paper, we focused on the iterative cycle in which behavior shapes the evidence people encounter, which shapes their beliefs, which in turn shapes the behavior etc. This cycle amplifies behavior if people feel (implicitly or explicitly) that their behavior produces desirable outcomes and we argued that this could lead to persisting biases. Importantly, this analysis may also offer some important insights for those aiming to change behavior.



First of all, it is important to realize that the biases we discussed here do not merely exist in the heads of people. As some have pointed out before (Fiedler, 2000; Fiedler & Wänke, 2009), biases can also be present in the evidence we encounter in the world, for example in the form of highly skewed evidence. By regarding information sampling as a process that often piggybacks on goal pursuit, we link the two together, with goal pursuit driving the bias in the evidence that is encountered. Because biased beliefs are based on this biased evidence, then, they reflect the subjective evidence people encounter. Therefore, telling people they are biased may be counterproductive and only create resistance. People's subjective beliefs and biases may very well be an accurate reflection of the evidence they encountered – regardless of whether an objective bias is present or not. These biases may therefore be hard to change, as they are based on first-hand experiences and people "saw it with their own eyes".

Yet this realization may also offer opportunities for interventions if they target the subjective experiences of people. Based on our framework, there are three ways in which the cycle may be broken. As noted earlier, changing people's behavior will change the evidence they encounter. This may push people toward other attractive options, which may lead to new cycles. Such changes may be powerful, as they (indirectly) change the evidence people encounter themselves. Targeting people's behavior directly targets their subjective experience.

A second approach could be to target people's beliefs by persuasion. However, this approach will only create lasting change if the resulting behavior also delivers the desirable outcomes that are necessary to keep the cycle going – a

critical step which interventions should heed. Otherwise, persuasion may fail as it does not lead to an iterative process and the arguments used are not backed up by real-world evidence. Reminiscent of the intention-behavior gap (Sheeran & Webb, 2016), targeting people's beliefs by persuasion may not be successful in and of itself but requires downstream behavior changes.

Finally, evidence that is encountered could be targeted, for instance, by attaching additional rewards to desirable outcomes. While this may create rapid change, the resulting behavior will only last as long as the rewards are present – unless this temporary intervention has led to experiences that confirmed the superiority of the targeted option. Without these rewards, the new behavior may be less attractive than before (Lepper et al., 1973), and old behaviors may become dominant again. This final type of interventions would have to critically assess whether a sufficient shift in the subjective experiences have taken place such that the rewards could be changed again.

To summarize, targeting behavior can be a powerful intervention because it directly affects the evidence one would encounter. On the other hand, if behavioral changes don't have any effects on the subjective experience, they, too, will be ineffective. Persuasion or manipulations of the feedback encountered can be equally effective. But, once again, the critical assessment must be whether they lead to changes in the subjective experience. In conclusion, interventions, and especially their critical appraisal, should not be restricted to any one stage but should be assessed within the iterative framework we present here.

It should be noticed that while we mainly focus on positive experiences and rewards, negative experiences or losses could also have a pervasive influence on our behavior. The hot-stove effect, demonstrated by Denrell and March (2001), holds that people will not revisit a negative option and thus do not obtain further evidence that may reveal the option is more positive than the first encounter did suggest. Such negative options may then create biases as long as there are other options (cf. Fetchenhauer & Dunning, 2010; Jaeger et al., 2021). And, in essence, exposure therapy (Bohnlein et al., 2020; Craske et al., 2014) can be seen as an extreme forced-sampling manipulation, aimed at creating benign firsthand experiences that may change the patient's beliefs. It is therefore reasonable to assume that our cyclic framework will generalize to sampling strategies given that they result in either seeking out or avoiding certain choice alternatives over others. When all options are poor, however, that is, no one option results in sufficient positive outcomes, people will eventually end up sampling in a fairly unbiased manner, as the negative outcomes promoted exploration of all possible alternatives (Harris, Fiedler, et al., 2022; Harris et al., 2020).

Finally, while our iterative cycle seems to be perpetuating itself infinitely, we would like to note that in real life, the values of choice options are usually not constant, but may change over time. Your favorite restaurant may get a new, less competent chef, your favorite venue another DJ, or you might be fed up with always eating the same sandwiches at your favorite sandwich place. As such, exploration may remain a potentially valuable strategy, even when all information is known (Hotaling et al., 2020; Navarro et al., 2016; Speekenbrink & Konstantinidis, 2015). We recognize that in dynamic environments, biases may deteriorate more quickly if people revisit options that may have changed over time (Biele et al., 2009). This tendency to explore may also be related to individual differences (Mehlhorn et al., 2015), causing some people to explore less and be more susceptible to the processes described by our iterative cycle. On the other hand, momentary changes in value could also create initial biases to begin with that are then upheld if people don't realize the environment is more dynamic. As such, changes in the value of options could work both for and against the processes we describe here.



This notion might be of particular interest to the literature on habits. Habit formation is often described as a process in which the behavior becomes evoked by the context in which it repeatedly is performed (see Verplanken, 2018; Wood & Rünger, 2016). Although rewards are considered to initially reinforce the behavior in the context, their role is reduced over time, so that eventually the behavior is fully determined by the context even when the behavior is no longer rewarding. The iterative cycle we describe here, however, suggests a different possibility: It may be the case that behavior is still performed in the same context because people's idiosyncratic experiences still maintain the representation of the behavior as rewarding – or at least as more rewarding than alternatives. Although people may keep performing behaviors that are, in fact, suboptimal compared to other alternatives, this does not mean that they are habitual in the sense that the behavior is fully driven by the context (Marien et al., 2019).

Consider, for instance, habitual car use. Although taking the car to work each morning (context) may have been rewarding in the past (fast and easy), people may stick to this behavior despite declining rewards (long traffic jams, reduced parking opportunities), not because the context evokes the behavior, but because people do not represent other alternatives (e.g., taking the train) as more rewarding, simply because they have not engaged in that behavior recently. As long as the "habitual" option is still idiosyncratically represented as the best option out there, people may not be inclined to explore unknown options that may yield more rewarding outcomes. As such, we predict that "habitual" behaviors are more common in reward-rich environments in which frequent positive outcomes maintain the cycle,

than in reward-impoverished environments in which frequent negative outcomes might push decision-makers to explore alternatives. According to this explanation, “habitual” behaviors may still be driven by anticipated rewards and not be directly evoked by the context. Our framework provides an explanation of how idiosyncratic representations of (relative) rewards can become biased by the very fact that the same behavioral option is chosen in the same context, which constrains learning about other alternatives. This alterative view on habits may offer new insights and avenues for future research and interventions in this area.

To conclude, by the choices people make they shape the evidence they encounter, which feeds into their beliefs, shaping behavior again, etc. With our choices we create an idiosyncratic experience and, ultimately, shape the world we perceive. This may lead to a biased representation of actions and outcomes, that in turn may lead to suboptimal behavior. To help people overcome such biases and to create lasting behavioral change, the iterative cycle can be broken at three points, providing options for interventions. Viewing people’s goal pursuit as a special case of evidence sampling, then, may be a perspective that allows for understanding and changing people’s everyday behavior.

# Appendices

## Appendix A: Supplementary Section for Chapter 2

Detailed results from Bayesian analyses

| Exp. | DV | Test | BF | 95% Highest Density Interval (HDI) | Median of 95% HDI | Robustness interval ($BF > 3$) |
|---|---|---|---|---|---|---|
| 1 | Sampling | Rich > .5 | $BF_{+0} = 5.34$ | [2.27, 3.54] | $d = 2.89$ | [0.02, 1.50] |
| | | Imp. = .5 | $BF_{01} = 1.49$ | [2.35, 3.72] | $d = 3.03$ | *NA* |
| | | Rich > Imp. | $BF_{+0} = 0.46$ | [-0.22, 0.52] | $d = 0.15$ | *NA* |
| | Oscillation | Imp. > Rich | $BF_{+0} = 0.24$ | [-0.20, 0.34] | $d = 0.07$ | *NA* |
| | Rel. cont. estimates | Rich > .5 | $BF_{+0} = 0.96$ | [-0.06, 0.48] | $d = 0.21$ | *NA* |
| | | Imp. = .5 | $BF_{01} = 6.36$ | [-0.26, 0.29] | $d = 0.01$ | [0.30, 1.50] |
| | | Rich > Imp. | $BF_{+0} = .63$ | [-0.17, 0.58] | $d = 0.20$ | *NA* |
| | Cond. estimates | Rich > .5 | $BF_{+0} = 10.05$ | [0.10, 0.65] | $d = 0.37$ | [0.03, 1.50] |
| | | Imp. = .5 | $BF_{01} = 6.37$ | [-0.26, 0.28] | $d = 0.01$ | [0.30, 1.50] |
| | | Rich > Imp. | $BF_{+0} = 2.95$ | [0.01, 0.77] | $d = 0.38$ | [0.11, 0.68] |
| | Confidence estimates | Freq. = Infreq. | $BF_{01} = 6.35$ | [-0.24, 0.30] | $d = 0.03$ | [0.30, 1.50] |
| 2a | Rel. cont. estimates - pre | Rich = Imp | $BF_{01} = 3.47$ | [-0.11, 0.42] | $d = 0.15$ | [0.61, 1.50] |
| | | All > 0 | $BF_{+0} = 411159.82$ | [0.25, 0.54] | $d = 0.40$ | [0.02, 1.50] |
| | Cond. Estimates – pre | Rich = Imp | $BF_{01} = 5.93$ | [-0.33, 0.21] | $d = -0.06$ | [0.32, 1.50] |
| | | All > 0 | $BF_{+0} = 16802.19$ | [0.21, 0.49] | $d = 0.35$ | [0.02, 1.50] |
| | Confidence - pre | Freq. = Infreq. | $BF_{01} = 8.31$ | [-0.15, 0.24] | $d = 0.04$ | [0.23, 1.50] |
| | Sampling | Rich > .5 | $BF_{+0} = 124.08$ | [1.81, 2.53] | $d = 2.17$ | [0.02, 1.50] |
| | | Imp. = .5 | $BF_{01} = 5.99$ | [2.15, 2.98] | $d = 2.56$ | [0.33, 1.50] |
| | | Rich > Imp. | $BF_{+0} = 5.53$ | [0.06, 0.60] | $d = 0.33$ | [0.03, 1.50] |
| | Oscillation | Imp. > Rich | $BF_{+0} > 1000000$ | [0.59, 1.24] | $d = 0.91$ | [0.02, 1.50] |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Rel. cont. estimate - post | Rich > .5 | $BF_{+0}$ = 9.82 | [0.08, 0.47] | $d$ = 0.27 | [0.02, 1.50] |
| | | Imp. = .5 | $BF_{01}$ = 3.78 | [-0.33, 0.07] | $d$ = 0.13 | [0.55, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 0.59 | [-0.10, 0.44] | $d$ = 0.17 | *NA* |
| | Cond. Estimates - post | Rich > .5 | $BF_{+0}$ = 2.14 | [0.02, 0.40] | $d$ = 0.21 | [0.05, 0.44] |
| | | Imp. = .5 | $BF_{01}$ = 7.74 | [-0.24, 0.14] | $d$ = -0.05 | [0.24, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 2.17 | [0.00, 0.54] | $d$ = 0.27 | [0.09, 0.39] |
| | Confidence - post | Freq. = Infreq. | $BF_{01}$ = 7.37 | [-0.25, 0.13] | $d$ = -0.06 | [0.26, 1.50] |
| 2b | Rel. cont. estimates - pre | Rich = Imp | $BF_{01}$ = 5.28 | [-0.18, 0.36] | $d$ = 0.09 | [0.36, 1.50] |
| | | All > 0 | $BF_{+0}$ > 1000000 | [0.27, 0.56] | $d$ = 0.42 | [0.02, 1.50] |
| | Base rates | Rich = Imp | $BF_{01}$ = 6.22 | [-0.31, 0.23] | $d$ = -0.04 | [0.30, 1.50] |
| | | All > 0 | $BF_{+0}$ > 1000000 | [0.53, 0.84] | $d$ = 0.69 | [0.02, 1.50] |
| | Cond. Estimates – pre | Rich = Imp | $BF_{01}$ = 6.45 | [-0.28, 0.26] | $d$ = -0.01 | [0.30, 1.50] |
| | | All > 0 | $BF_{+0}$ = 1000000 | [0.34, 0.64] | $d$ = 0.49 | [0.02, 1.50] |
| | Confidence - pre | Freq. = Infreq. | $BF_{01}$ = 3.69 | [-0.06, 0.32] | $d$ = 0.13 | [0.56, 1.50] |
| | Sampling | Rich > .5 | $BF_{+0}$ = 8.56 | [1.92, 2.69] | $d$ = 2.17 | [0.02, 1.50] |
| | | Imp. = .5 | $BF_{01}$ = 0.04 | [2.46, 3.37] | $d$ = 2.92 | *NA* |
| | | Rich > Imp. | $BF_{+0}$ = 0.26 | [-0.19, 0.34] | $d$ = 0.08 | *NA* |
| | Oscillation | Imp. > Rich | $BF_{+0}$ > 1000000 | [0.72, 1.38] | $d$ = 1.05 | [0.02, 1.50] |
| | Rel. cont. estimate - post | Rich > .5 | $BF_{+0}$ = 4.65 | [0.05, 0.45] | $d$ = 0.25 | [0.03, 1.18] |
| | | Imp. = .5 | $BF_{01}$ = 2.23 | [-0.03, 0.36] | $d$ = 0.17 | [0.98, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 0.36 | [-0.15, 0.39] | $d$ = 0.12 | [0.79, 1.50] |
| | Base rates | Rich > .5 | $BF_{+0}$ = 12.78 | [0.09, 0.49] | $d$ = 0.29 | [0.02, 1.50] |

A

| | | | | | |
|---|---|---|---|---|---|
| | Imp. = .5 | $BF_{01}$ = 1.02 | [-0.40, -0.01] | $d$ = -0.21 | *NA* |
| | Rich > Imp. | $BF_{+0}$ = 0.61 | [-0.10, 0.44] | $d$ = 0.17 | *NA* |
| Cond. Estimates - post | Rich > .5 | $BF_{+0}$ = 4.09 | [0.04, 0.44] | $d$ = 0.24 | [0.03, 1.02] |
| | Imp. = .5 | $BF_{01}$ = 6.64 | [-0.12, 0.27] | $d$ = -0.08 | [0.29, 1.50] |
| | Rich > Imp. | $BF_{+0}$ = 1.03 | [-0.05, 0.49] | $d$ = 0.22 | *NA* |
| Confidence - post | Freq. = Infreq | $BF_{01}$ = 7.35 | [-0.13, 0.25] | $d$ = -0.06 | [0.26, 1.50] |

Robustness intervals are calculated by testing which BFs are larger than $BF = 3$ for prior Cauchy distributions with a width of $0 < r \leq 1.5$.

## Appendix B: Supplementary Section for Chapter 3

### Supplementary Results

#### *Experiment 1*

**Predictability and Control.** The estimates for predictability and control confirmed the pattern on the other dependent measures. For predictability, we found a bias in the reward-rich condition though the Bayesian analysis was inconclusive, $\Delta P_{\text{rich}} = .04$ ($SD = 0.22$), $BF_{+0} = 1.16$, $t(100) = 1.88$, $p = .031$, $d = 0.19$, 95% CI [-0.01, 0.38]. As expected, participants in the reward-impoverished condition did not differ from zero, $\Delta P_{\text{impoverished}} = .00$ ($SD = 0.20$), $BF_{01} = 9.02$, $t(99) = 0.06$, $p = .951$, $d = 0.01$, 95% CI [-0.19, 0.20]. There was neither a difference between conditions ($BF_{+0} = 0.65$, $t(197.38) = 1.35$, $p = .089$, $d = 0.19$, 95% CI [-0.09, 0.47]) nor an overall effect, $BF_{+0} = 0.40$, $t(200) = 1.44$, $p = .076$, $d = 0.10$, 95% CI [-0.04, 0.24].

On the control measure, too, participants in the reward-rich condition did not exhibit a bias, $\Delta P_{\text{rich}} = .04$ ($SD = 0.25$), $BF_{+0} = 0.66$, $t(100) = 1.55$, $p = .062$, $d = 0.15$, 95% CI [-0.04, 0.35]. Participants in the reward-impoverished condition did not differ from zero, $\Delta P_{\text{impoverished}} = .00$ ($SD = 0.18$), $BF_{01} = 8.73$, $t(99) = 0.26$, $p = .792$, $d = 0.03$, 95% CI [-0.17, 0.22]. And, again, there was neither a difference between conditions ($BF_{+0} = 0.47$, $t(178.83) = 1.12$, $p = .132$, $d = 0.16$, 95% CI [-0.12, 0.44]) nor an overall effect, $BF_{+0} = 0.39$, $t(200) = 1.43$, $p = .078$, $d = 0.10$, 95% CI [-0.04, 0.24].

#### *Experiment 2*

**Predictability and Control Post-Measures.** We found no evidence for biases in the reward-rich condition in how predictable they estimated the task to be, $\Delta P_{\text{rich}} = .02$ ($SD = 0.21$), $BF_{+0} = 0.27$, $t(97) = 0.91$, $p = .182$, $d = 0.09$, 95% CI [-0.11, 0.29]. There was also no indication of any bias in the reward-impoverished condition, $\Delta P_{\text{impoverished}} = .00$ ($SD = 0.11$), $BF_{01} = 8.96$, $t(100) = -0.16$, $p = .873$, $d = -0.02$, 95% CI [-0.21, 0.18]. Accordingly, we found neither a difference between conditions ($BF_{+0} = 0.36$, $t(148.29) = 0.88$, $p = .190$, $d = 0.13$, 95% CI [-0.15, 0.41]), nor an overall effect, $BF_{+0} = 0.16$, $t(198) = 0.72$, $p = .235$, $d = 0.05$, 95% CI [-0.09, 0.19].

This same pattern held true in how much control participants perceived. There was no systematic bias in the reward-rich condition ($\Delta P_{\text{rich}} = .04$, $SD = 0.23$, $BF_{+0} = 0.79$, $t(97) = 1.66$, $p = .050$, $d = 0.17$, 95% CI [-0.03, 0.37]) nor in the reward-impoverished condition, $\Delta P_{\text{impoverished}} = .01$ ($SD = 0.12$), $BF_{01} = 7.14$, $t(100) = 0.70$, $p = .484$, $d = 0.07$, 95% CI [-0.13, 0.27]. Therefore there was also no difference between conditions, $BF_{+0} = 0.49$, $t(148.31) = 1.13$, $p = .130$, $d = 0.16$, 95% CI [-0.12, 0.44]. There was, however, a small overall effect, $BF_{+0} = 0.73$, $t(198) = 1.79$, $p = .037$, $d = 0.13$, 95% CI [-0.01, 0.27].

***Experiment 3***

**Predictability and Control.** We again calculated difference scores from the predictability and control estimates participants made. These indicated a bias of participants in the reward-rich condition, $\Delta P_{rich}$ = .14 (*SD* = 0.26), $BF_{+0}$ = 126.54, *t*(49) = 3.79, *p* < .001, *d* = 0.54, 95% CI [0.24, 0.83]. Participants in the reward-impoverished condition did not differ from zero, $\Delta P_{impoverished}$ = .00 (*SD* = 0.14), $BF_{01}$ = 6.47, *t*(49) = -0.10, *p* = .920, *d* = -0.01, 95% CI [-0.29, 0.26]. The two conditions differed from one another ($BF_{+0}$ = 60.39, *t*(74.3) = 3.40, *p* < .001, *d* = 0.68, 95% CI [0.27, 1.09]), but there was also an overall bias, $BF_{+0}$ = 21.70, *t*(99) = 3.14, *p* = .001, *d* = 0.31, 95% CI [0.11, 0.52].

We found this same pattern for the control measure. Participants in the reward-rich condition exhibited a bias, $\Delta P_{rich}$ = .13 (*SD* = 0.25), $BF_{+0}$ = 68.36, *t*(49) = 3.57, *p* < .001, *d* = 0.50, 95% CI [0.21, 0.80]. Participants in the reward-impoverished condition again did not differ from zero, $\Delta P_{impoverished}$ = -.01 (*SD* = 0.18), $BF_{01}$ = 6.08, *t*(49) = -0.37, *p* = .711, *d* = -0.05, 95% CI [-0.33, 0.22]. The two conditions differed from one another ($BF_{+0}$ = 29.57, *t*(87.67) = 3.14, *p* = .001, *d* = 0.63, 95% CI [0.22, 1.03]), and there was also an overall bias, $BF_{+0}$ = 5.28, *t*(99) = 2.60, *p* = .005, *d* = 0.26, 95% CI [0.06, 0.46].

**Detailed Results From Bayesian Analyses**

| Exp. | DV | Test | BF | 95% Highest Density Interval (HDI) | Median of 95% HDI | Robustness interval (BF > 3) |
|---|---|---|---|---|---|---|
| 1 | Rel. contingency | Rich > .5 | $BF_{+0}$ = 1.02 | [-0.02, 0.37] | *d* = 0.17 | *NA* |
| | | Imp. = .5 | $BF_{01}$ = 8.92 | [-0.21, 0.17] | *d* = -0.02 | [0.21, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 0.71 | [-0.08, 0.46] | *d* = 0.19 | *NA* |
| | | Overall | $BF_{+0}$ = 0.28 | [-0.05, 0.22] | *d* = 0.08 | *NA* |
| | Cond. estimates | Rich > .5 | $BF_{+0}$ = 0.73 | [-0.04, 0.35] | *d* = 0.16 | *NA* |
| | | Imp. = .5 | $BF_{01}$ = 4.94 | [-0.09, 0.30] | *d* = 0.11 | [0.41, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 0.21 | [-0.22, 0.31] | *d* = 0.05 | *NA* |
| | | Overall | $BF_{+0}$ = 0.97 | [0.00, 0.27] | *d* = 0.14 | *NA* |
| | Confidence | | $BF_{01}$ = 4.08 | [-0.07, 0.31] | *d* = 0.12 | [0.50, 1.50] |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | Rel. contingency pre | Rich > .5 | $BF_{+0}$ = 59,742.96 | [0.33, 0.75] | $d$ = 0.53 | [0.02, 1.50] |
| | | Imp. < .5 | $BF_{-0}$ = 0.03 | [0.08, 0.48] | $d$ = 0.28 | *NA* |
| | | Rich = Imp. | $BF_{01}$ = 1.86 | [-0.06, 0.48] | $d$ = 0.22 | [1.24, 1.50] |
| | | Overall | $BF_{+0}$ = 591,207.92 | [0.26, 0.55] | $d$ = 0.40 | [0.02, 1.50] |
| | Cond. Estimates pre | Rich > .5 | $BF_{+0}$ = 49,082.87 | [0.32, 0.74] | $d$ = 0.53 | [0.02, 1.50] |
| | | Imp. < .5 | $BF\text{-}_{0}$ = 0.05 | [-0.05, 0.34] | $d$ = 0.14 | *NA* |
| | | Rich > Imp. | $BF_{01}$ = 0.31 | [0.08, 0.63] | $d$ = 0.34 | *NA* |
| | | Overall | $BF_{+0}$ = 3,699.25 | [0.18, 0.47] | $d$ = 0.32 | [0.02, 1.50] |
| | Confidence pre | | $BF_{01}$ = 1.99 | [-0.02, 0.36] | $d$ = 0.17 | [1.12, 1.50] |
| | Rel. contingency | Rich > .5 | $BF_{+0}$ = 1.97 | [0.01, 0.40] | $d$ = 0.21 | [0.06, 0.38] |
| | | Imp. = .5 | $BF_{01}$ = 0.53 | [0.04, 0.43] | $d$ = 0.24 | *NA* |
| | | Rich > Imp. | $BF_{+0}$ = 0.15 | [-0.27, 0.26] | $d$ = -0.01 | *NA* |
| | | Overall | $BF_{+0}$ = 25.26 | [0.09, 0.37] | $d$ = 0.23 | [0.02, 1.50] |
| | Cond. estimates | Rich > .5 | $BF_{+0}$ = 3.01 | [0.03, 0.43] | $d$ = 0.23 | [0.03, 0.71] |
| | | Imp. = .5 | $BF_{01}$ = 0.24 | [0.07, 0.47] | $d$ = 0.27 | *NA* |
| | | Rich > Imp. | $BF_{+0}$ = 0.15 | [-0.27, 0.26] | $d$ = -0.01 | *NA* |
| | | Overall | $BF_{+0}$ = 82.35 | [0.11, 0.39] | $d$ = 0.25 | [0.02, 1.50] |
| | Confidence | | $BF_{01}$ = 7.37 | [-0.14, 0.25] | $d$ = 0.06 | [0.26, 1.50] |
| 3 | Rel. contingency | Rich > .5 | $BF_{+0}$ = 71.00 | [0.19, 0.77] | $d$ = 0.48 | [0.02, 1.50] |
| | | Imp. = .5 | $BF_{01}$ = 4.31 | [-0.39, 0.15] | $d$ = -0.12 | [0.47, 1.50] |
| | | Rich > Imp. | $BF_{+0}$ = 1.57 | [-0.07, 0.69] | $d$ = 0.31 | *NA* |
| | | Overall | $BF_{+0}$ = 0.82 | [-0.03, 0.36] | $d$ = 0.16 | *NA* |

A

| | | | | | |
|---|---|---|---|---|---|
| Cond. estimates | Rich > .5 | $BF_{+0}$ = 16.25 | [0.11, 0.69] | $d$ = 0.40 | [0.02, 1.50] |
| | Imp. = .5 | $BF_{01}$ = 1.97 | [-0.49, 0.06] | $d$ = -0.21 | [1.17, 1.50] |
| | Rich > Imp. | $BF_{+0}$ = 0.66 | [-0.16, 0.58] | $d$ = 0.20 | *NA* |
| | Overall | $BF_{+0}$ = 0.34 | [-0.09, 0.30] | $d$ = 0.11 | *NA* |
| Confidence | Freq. = Infreq | $BF_{01}$ = 6.47 | [-0.25, 0.28] | $d$ = 0.01 | [0.29, 1.50] |

## Appendix C: Supplementary Section for Chapter 4

Detailed results from Bayesian analyses

| DV | Test | BF | 95% Highest Density Interval (HDI) | Median of 95% HDI | Robustness interval ($BF$ > 3) |
|---|---|---|---|---|---|
| Rel. cont. estimates - pre | PC > 0 | $BF_{+0}$ = 0.11 | [-0.19, 0.19] | $d$ = 0.00 | *NA* |
| | DN < 0 | $BF_{-0}$ = 0.05 | [-0.07, 0.32] | $d$ = 0.12 | *NA* |
| Cond. Estimates – pre | PC > 0 | $BF_{+0}$ = 0.28 | [-0.10, 0.29] | $d$ = 0.09 | *NA* |
| | DN < 0 | $BF_{-0}$ = 0.05 | [-0.08, 0.31] | $d$ = 0.12 | *NA* |
| Confidence - pre | Freq. = Infreq | $BF_{01}$ = 4.14 | [-0.14, 0.40] | $d$ = 0.13 | [0.48, 1.50] |
| Sampling | PC > 0 | $BF_{+0}$ = 0.07 | [0.00, 0.18] | $d$ = 0.05 | *NA* |
| | DN < 0 | $BF_{-0}$ = 0.68 | [-0.35, -0.02] | $d$ = -0.16 | *NA* |
| Rel. cont. estimate - post | PC > 0 | $BF_{01}$ = 8.88 | [-0.17, 0.21] | $d$ = 0.02 | [0.21, 1.50] |
| | DN < 0 | $BF_{-0}$ = 0.08 | [-0.16, 0.23] | $d$ = 0.04 | *NA* |
| Cond. Estimates - post | PC > 0 | $BF_{01}$ = 8.13 | [-0.15, 0.23] | $d$ = 0.04 | [0.23, 1.50] |
| | DN < 0 | $BF_{-0}$ = 1.13 | [-0.38, 0.01] | $d$ = -0.18 | *NA* |
| Confidence - post | Freq. = Infreq | $BF_{01}$ = 6.46 | [-0.27, 0.27] | $d$ = 0.01 | [0.30, 1.50] |

## Appendix D: Supplementary Section for Chapter 5

### Supplementary Methods and Results

#### *Methods – Data Preparation*

Data preparation and analyses were undertaken using R (R Core Team, 2018) and especially the packages "BayesFactor" (Morey & Rouder, 2018), "dplyr" (Wickham et al., 2019), "lme4" (Bates et al., 2015), and "lmerTest" (Kuznetsova et al., 2017). All Bayesian tests use the default prior of the Bayes-Factor package, namely a Cauchy distribution of width $r = \frac{\sqrt{2}}{2}$.

#### *Supplemental Results for Experiment 1*

**Pre-Measures.** Following the induction phase, participants indicated their relative contingency estimate as well as conditional probability estimates. The objectively better strategy would be to choose the infrequent option, while a successful pseudocontingency-induction would result in a preference for the frequent option. With an average preference of $M = 5.28$ ($SD = 32.52$), the pseudocontingency illusion clearly overruled the genuine contingency. There was no support for participants favoring the objectively better option, $BF_{-0} = 0.04$, $t(99) = 1.62$, $p = .946$, $d = 0.16$, 95% CI [-0.03, 0.36]. However, the reversal was not to the extent that the pseudocontingency illusion would become statistically significant, $BF_{+0} = 0.74$, $t(99) = 1.62$, $p = .054$, $d = 0.16$, 95% CI [-0.03, 0.36].

Similarly, participants indicated winning with the frequent bag 50.24% of the time ($SD = 27.84$) and 40.28% of the time ($SD = 26.76$) with the infrequent bag. The mean $\Delta P$-score was $\Delta P = .06$ ($SD = 0.37$), which, again, indicates that participants clearly did not favor the objectively better option, $BF_{-0} = 0.04$, $t(99) = 1.74$, $p = .957$, $d = 0.17$, 95% CI [-0.02, 0.37]. Instead, the pseudocontingency even overruled the genuine contingency, $BF_{+0} = 0.90$, $t(99) = 1.74$, $p = .043$, $d = 0.17$, 95% CI [-0.02, 0.37].

**Confidence**. Overall, participants indicated that they were more confident in their estimates regarding the frequent bag relative to the infrequent bag ($\Delta P = .10$, $SD = 0.26$), $BF_{10} = 88.27$, $t(99) = 3.83$, $p < .001$, $d = 0.38$, 95% CI [0.18, 0.59]. This preference was mainly due to the biased group that indicated a stronger difference in their confidence between bags ($\Delta P = .17$, $SD = 0.27$, $BF_{10} = 604.95$, $t(51) = 4.54$, $p < .001$, $d = 0.63$, 95% CI [0.33, 0.93]), while there was no such difference in the unbiased group ($\Delta P = .03$, $SD = 0.23$), $BF_{01} = 4.80$, $t(47) = 0.78$, $p = .442$, $d = 0.11$, 95% CI [-0.17, 0.40].

**Sampling.** Following these initial estimates, participants engaged in 83 trials in which they could freely choose between both options while trying to earn points. Choices on the first trial largely matched people's indicated preferences in

the first estimation phase with 42/52 choosing the frequent option in bias group but only 20/48 in the no-bias group.

Across all trials one might expect participants in the bias group to attenuate their initial biases and to choose the objectively better option more frequently. However there was no support for this hypothesis ($M$ = 60.91, $SD$ = 27.94), $BF_{-0}$ = 0.04, $t$(51) = 2.82, $p$ = .997, $d$ = 0.39, 95% CI [0.11, 0.67]. Instead, averaged across trials, they sampled the initially frequent option more, $BF_{+0}$ = 10.17, $t$(51) = 2.82, $p$ = .003, $d$ = 0.39, 95% CI [0.11, 0.67]. Only participants in the no-bias group selected the initially more infrequent (but actually better in terms of payoffs) bag more often ($M$ = 39.22, $SD$ = 24.51), $BF_{10}$ = 6.85, $t$(47) = -2.94, $p$ = .005, $d$ = -0.42, 95% CI [-0.72, -0.13].

**Post-Measures**. After the sampling phase, in the light of all 100 observations, participants again indicated their relative contingency estimate as well as conditional probability estimates. The biased group still did not report preferences for the objectively better option ($M$ = 0.88, $SD$ = 35.11), $BF_{-0}$ = 0.13, $t$(51) = 0.18, $p$ = .572, $d$ = 0.03, 95% CI [-0.25, 0.30]. Next, we tested whether the pseudocontingency illusion would still fully reverse the genuine contingency as the positive mean might indicate. That however, was no longer the case, $BF_{+0}$ = 0.17, $t$(51) = 0.18, $p$ = .428, $d$ = 0.03, 95% CI [-0.25, 0.30]. The unbiased group, on the other hand, indicated a clear preference for the objectively better option ($M$ = -16.58, $SD$ = 32.09), $BF_{-0}$ = 69.62, $t$(47) = -3.58, $p$ < .001, $d$ = -0.52, 95% CI [-0.82, -0.22].

Participants in the biased group indicated to have won 64.05% ($SD$ = 26.82) of the time with the frequent bag and 54.90% ($SD$ = 27.14) of the time with the infrequent bag. The resulting $\Delta P$-score was $\Delta P$ = .09 ($SD$ = 0.44), again pointed in the direction of a pseudocontingency illusion overruling the objective contingency, $BF_{-0}$ = 0.06, $t$(51) = 1.48, $p$ = .928, $d$ = 0.21, 95% CI [-0.07, 0.48]. However, as with the relative contingency estimate, the pseudocontingency illusion no longer fully reversed the genuine contingency, $BF_{+0}$ = 0.78, $t$(51) = 1.48, $p$ = .072, $d$ = 0.21, 95% CI [-0.07, 0.48].

Participants in the no-bias group indicated to have won 53.00% ($SD$ = 27.05) of the time with the frequent bag and 66.88% ($SD$ = 24.78) of the time with the infrequent bag. In other words, they correctly identified the infrequent bag to result in more wins, $\Delta P$ = -.14 ($SD$ = 0.40), $BF_{-0}$ = 3.97, $t$(47) = -2.38, $p$ = .011, $d$ = -0.34, 95% CI [-0.64, -0.05].

**Confidence**. Participants in the bias group indicated being slightly more confident in the frequent bag ($\Delta P$ = .07, $SD$ = 0.34). This difference, however, was not statistically significant, $BF_{01}$ = 2.12, $t$(51) = 1.56, $p$ = .124, $d$ = 0.22, 95% CI [-0.06, 0.49]. In the no-bias group, on the other hand, participants indicated being

slightly more confident in the infrequent bag ($\Delta P$ = -.09, $SD$ = .30), $BF_{01}$ = 1.35, $t$(47) = -2.18, $p$ = .034, $d$ = -0.31, 95% CI [-0.60, -0.02].

***Supplemental Results for Experiment 2***

**Pre-Measures**. Following the induction phase, participants again indicated their relative contingency estimate as well as conditional probability estimates. Across all participants the pseudocontingency illusion still suppressed the genuine contingency. The data did not indicate a preference towards the objectively better option ($M$ = -0.64, $SD$ = 31.30), $BF_{-0}$ = 0.13, $t$(99) = -0.20, $p$ = .419, $d$ = -0.02, 95% CI [-0.22, 0.18]. However, as already indicated by the negative mean, the pseudocontingency illusion did not overrule the genuine contingency.

Furthermore, participants indicated winning with the frequent bag 58.63% of the time ($SD$ = 19.68) and 57.31% of the time ($SD$ = 21.68) with the infrequent bag. The mean $\Delta P$-score then was $\Delta P$ = .01 ($SD$ = 0.32), which is again reflective of the failure to figure out the objectively better option, $BF_{-0}$ = 0.08, $t$(99) = 0.41, $p$ = .659, $d$ = 0.04, 95% CI [-0.16, 0.24]. While the mean $\Delta P$ was positive, the pseudocontingency illusion did not override the genuine contingency to the extent that we would find a significant effect, $BF_{+0}$ = 0.16, $t$(99) = 0.41, $p$ = .341, $d$ = 0.04, 95% CI [-0.16, 0.24].

We then again created a dummy variable for participants based on the estimates from both measures. Again, we grouped all participants whose conditional probability estimates reflected a pseudocontingency (i.e., an above-zero $\Delta P$ difference) in the bias group ($n_{bias}$ = 42). And we grouped all participants that indicated no initial bias in either of their estimates or that made estimates in the objectively correct direction ($n_{nobias}$ = 57).

**Confidence**. Overall, participants indicated that they were equally confident in their estimates for both bags ($\Delta P$ = .01, $SD$ = 0.18), $BF_{01}$ = 7.12, $t$(99) = 0.70, $p$ = .486, $d$ = 0.07, 95% CI [-0.13, 0.27]. Further analyses revealed that the biased group indicated a slightly stronger confidence in the frequent bag ($\Delta P$ = .06, $SD$ = 0.18, $BF_{10}$ = 1.17, $t$(42) = 2.08, $p$ = .043, $d$ = 0.32, 95% CI [0.01, 0.62]) relative to the unbiased group ($\Delta P$ = -.02, $SD$ = 0.17), $BF_{01}$ = 4.38, $t$(56) = -0.98, $p$ = .331, $d$ = -0.13, 95% CI [-0.39, 0.13]. Neither $\Delta P$ difference, however, was significant.

**Sampling**. Following these initial estimates, participants engaged in 83 trials in which they could freely choose between both options. Choices on the first trial again largely matched people's indicated preferences in the first estimation phase with 37/43 choosing the frequent option in the bias group, in contrast to only 20/57 in the no-bias group.

Again, participants' choices in the bias group did not indicate sufficient attenuation of their initial bias ($M$ = 52.73, $SD$ = 25.68), $BF_{-0}$ = 0.10, $t$(42) = 0.70, $p$ = .755, $d$ = 0.11, 95% CI [-0.19, 0.41]. Neither, however, did their choices suggest a

consistent bias for the initially frequent option, $BF_{+0}$ = 0.31, $t(42) = 0.70$, $p = .245$, $d = 0.11$, 95% CI [-0.19, 0.41]. The no-bias group in contrast selected the initially more infrequent (but actually better in terms of payoffs) bag more often ($M = 31.14$, $SD = 24.16$), $BF_{10}$ = 64349.60, $t(56) = -5.89$, $p < .001$, $d = -0.78$, 95% CI [-1.08, -0.48].

**Post-Measures**. After the sampling phase, participants again indicated their relative contingency estimate as well as conditional probability estimates. The biased group still did not report preferences for the objectively better option ($M = -0.37$, $SD = 36.54$), $BF_{-0}$ = 0.17, $t(42) = -0.07$, $p = .474$, $d = -0.01$, 95% CI [-0.31, 0.29]. The pseudocontingency illusion, however, did not overrule the genuine contingency. Instead, we found convincing evidence that participants' relative contingency estimates did not differ zero, $BF_{01}$ = 6.05, $t(42) = -0.07$, $p = .947$, $d = -0.01$, 95% CI [-0.31, 0.29]. In contrast, the unbiased group indicated a clear preference for the objectively better (initially infrequent) option ($M = -16.58$, $SD = 32.09$), $BF_{-0}$ = 77282.02, $t(56) = -5.75$, $p < .001$, $d = -0.76$, 95% CI [-1.06, -0.47].

For the conditional probability estimates participants in the biased group estimated to have won 56.40% ($SD = 23.87$) of the time with the frequent bag and 61.05% ($SD = 24.63$) of the time with the infrequent bag. The resulting $\Delta P$-score was $\Delta P = -.05$ ($SD = 0.41$), which went in the direction of the genuine contingency yet still not convincingly so, $BF_{-0} = 0.32$, $t(42) = -0.74$, $p = .232$, $d = -0.11$, 95% CI [-0.41, 0.19]. However, as with the relative contingency estimates the data most strongly supported there being no difference between options, $BF_{01} = 4.69$, $t(42) = -0.74$, $p = .232$, $d = -0.11$, 95% CI [-0.41, 0.19].

Participants in the no-bias group indicated to have won 51.86% ($SD = 21.19$) of the time with the frequent bag and 68.21% ($SD = 21.44$) of the time with the infrequent bag. In other words, they correctly identified the infrequent bag to result in more wins, $\Delta P = -.16$ ($SD = 0.33$), $BF_{-0} = 121.43$, $t(56) = -3.76$, $p < .001$, $d = -0.50$, 95% CI [-0.77, -0.22].

**Confidence**. Participants indicated being slightly more confident in the infrequent bag both in the biased group ($\Delta P = -.03$, $SD = 0.28$) as well as in the no-bias group ($\Delta P = -.03$, $SD = 0.23$). This difference, however, was not statistically significant, neither in the biased group ($BF_{01} = 4.77$, $t(42) = -0.71$, $p = .48$, $d = -0.11$, 95% CI [-0.41, 0.19]) nor in the not biased group ($BF_{01} = 4.63$, $t(56) = -0.92$, $p = .362$, $d = -0.12$, 95% CI [-0.38, 0.14]) suggesting overall similar confidence in the estimates participants made regarding the two options.

**Win-Stay-Lose-Shift Proof**

Assume that p(A) and p(B) describe the probabilities of choosing the two choice alternatives, A and B, respectively. Then assume that the probability of winning is:

$$p(win|A) = .8 \; and$$

$$p(win|B) = .75 \qquad (D.1)$$

and of losing:

$$p(loss|A) = 1 - p(win|A) = .2 \quad and$$

$$p(loss|B) = 1 - p(win|B) = .25 \qquad (D.2)$$

Then the chance for selecting option B on the next trial is the combined probability of having lost with A on the previous trial (shifting) and the probability of having won with B on the previous trial (staying). In other words:

$$p(B_t) = p(loss|A) * p(A_{t-1}) + p(win|B) * p(B_{t-1})$$

$$= .2 * p(A_{t-1}) + .75 * p(B_{t-1}) \qquad (D.3)$$

The probability of choosing A is the probability of not choosing B, so we can simplify to:

$$p(B_t) = .2 * [1 - p(B_{t-1})] + .75 * p(B_{t-1})$$

$$= .2 - .2 * p(B_{t-1}) + .75 * p(B_{t-1})$$

$$= .2 + .55 * p(B_{t-1}) \qquad (D.4)$$

Accordingly, the probability for choosing B at the first time point will be:

$$p(B_1) = .2 + .55 * p(B_0) \qquad (D.5)$$

The probability for choosing B at the second time point will be:

$$p(B_2) = .2 + .55 * p(B_1)$$

$$= .2 + .55 * [.2 + .55 * p(B_0)]$$

$$= .2 + .55 * .2 + .55 * .55 * p(B_0)$$

$$= .2 * (1 + .55) + .55^2 * p(B_0) \qquad (D.6)$$

A

And at the third time point:

$$p(B_3) = .2 + .55 * p(B_2)$$

$$= .2 + .55 * [.2 * (1 + .55) + .55^2 * p(B_0)]$$

$$= .2 + .55 * [.2 + .2 * .55 + .55^2 * p(B_0)]$$

$$= .2 + .55 * .2 + .55 * .55 * .2 + .55^3 * p(B_0)$$

$$= .2 * (1 + .55 + .55^2) + .55^3 * p(B_0) \qquad (D.7)$$

Generalizing this pattern, we can describe the probability of choosing B at time point *t* as:

$$p(B_t) = .2 * (1 + .55^1 + .55^2 + \ldots + .55^{t-2} + .55^{t-1}) + .55^t * p(B_0) \qquad (D.8)$$

Because:

$$\lim_{t\to\infty}(.55^t) = 0 \qquad (D.9)$$

We can conclude that:

$$\lim_{t\to\infty}\left(.55^t * \mathrm{p}(B_0)\right) = 0 \qquad (D.10)$$

Therefore:

$$\lim_{t\to\infty}\left(\mathrm{p}(B_{t-1})\right) = .2 * (1 + .55^1 + .55^2 + \ldots + .55^{t-2})$$

$$\lim_{t\to\infty}\left(\mathrm{p}(B_t)\right) = .2 * (1 + .55^1 + .55^2 + \ldots + .55^{t-2} + .55^{t-1})$$

$$= .2 * (1 + .55^1 + .55^2 + \ldots + .55^{t-2} + 0)$$

$$= .2 * (1 + .55^1 + .55^2 + \ldots + .55^{t-2})$$

$$= \lim_{t\to\infty}\left(\mathrm{p}(B_{t-1})\right) \qquad (D.11)$$

We can then conclude that for large *t*:

$$p(B_t) \approx p(B_{t-1}) \qquad (D.12)$$

Therefore, we can simplify the earlier line for large *t*:

$$p(B_t) = .2 + .55 * p(B_{t-1}) \qquad (D.13)$$

To:

$$p(B_t) = .2 + .55 * p(B_t)$$

$$.45 * p(B_t) = .2$$

$$p(B_t) = \frac{.2}{.45} \qquad (D.14)$$

In conclusion:

$$\lim_{t\to\infty}\left(\mathrm{p}(B_t)\right) \approx .444 \qquad (D.15)$$

For Experiment 2, we have

$$p(win|A) = .8 \; and$$

$$p(win|B) = .67$$

$$p(loss|A) = 1 - p(win|A) = .2$$

$$p(loss|B) = 1 - p(win|B) = .33 \qquad (D.16)$$

Then:

$$p(B_t) = .2 + .47 * p(B_{t-1}) \qquad (D.17)$$

And:

$$\lim_{t\to\infty}\left(\mathrm{p}(B_t)\right) \approx .377 \qquad (D.18)$$

A

**Detailed Results From Bayesian Analyses**

The following table includes more details on all Bayesian tests run in the paper and supplemental material. We report the Bayes Factors as well as 95% highest density intervals (HDI), the median of these intervals, and robustness intervals for which the Bayes Factors are larger than 3.

| Exp. | DV | Test | BF | 95% Highest Density Interval (HDI) | Median of 95% HDI | Robustness interval (BF > 3) |
|---|---|---|---|---|---|---|
| 1 | Rel. contingency pre | All < 0 | $BF_{-0}$ = 0.04 | [-0.04, 0.35] | *d* = 0.16 | *NA* |
| | | All > 0 | $BF_{+0}$ = 0.74 | [-0.04, 0.35] | *d* = 0.16 | *NA* |
| | Cond. Estimates pre | All < 0 | $BF_{-0}$ = 0.04 | [-0.03, 0.36] | *d* = 0.17 | *NA* |
| | | All > 0 | $BF_{+0}$ = 0.90 | [-0.03, 0.36] | *d* = 0.17 | *NA* |
| | Confidence pre | All ≠ 0 | $BF_{10}$ = 88.27 | [0.17, 0.57] | *d* = 0.37 | [0.02, 1.50] |
| | | Bias ≠ 0 | $BF_{10}$ = 604.95 | [0.31, 0.90] | *d* = 0.60 | [0.02, 1.50] |
| | | No-bias ≠ 0 | $BF_{01}$ = 4.80 | [-0.17, 0.37] | *d* = 0.10 | [0.41, 1.50] |
| | Sampling | All > 0.444 | $BF_{+0}$ = 1.85 | [1.43, 2.05] | *d* = 1.73 | [0.06, 0.33] |
| | | All ≠ 0.5 | $BF_{01}$ = 8.90 | [1.42, 2.05] | *d* = 1.73 | [0.21, 1.50] |
| | Rel. contingency post | Bias < 0 | $BF_{-0}$ = 0.13 | [-0.25, 0.28] | *d* = 0.02 | *NA* |
| | | Bias > 0 | $BF_{+0}$ = 0.17 | [-0.24, 0.29] | *d* = 0.02 | *NA* |
| | | No-bias < 0 | $BF_{-0}$ = 69.62 | [-0.79, -0.20] | *d* = -0.49 | [0.02, 1.50] |
| | Cond. Estimates post | Bias < 0 | $BF_{-0}$ = 0.06 | [-0.07, 0.46] | *d* = 0.19 | *NA* |
| | | Bias > 0 | $BF_{+0}$ = 0.78 | [-0.07, 0.46] | *d* = 0.19 | *NA* |
| | | No-bias < 0 | $BF_{-0}$ = 3.97 | [-0.61, -0.04] | *d* = -0.32 | [0.05, 1.03] |
| | Confidence post | Bias ≠ 0 | $BF_{01}$ = 2.12 | [-0.06, 0.47] | *d* = 0.20 | [1.08, 1.50] |
| | | No-bias ≠ 0 | $BF_{10}$ = 1.35 | [-0.58, -0.02] | *d* = -0.29 | *NA* |
| 2 | Rel. contingency pre | All < 0 | $BF_{-0}$ = 0.13 | [-0.21, 0.17] | *d* = -0.02 | *NA* |

| | | | | | |
|---|---|---|---|---|---|
| | All > 0 | $BF_{+0}$ = 0.09 | [-0.21, 0.18] | *d* = -0.02 | *NA* |
| Cond. Estimates pre | All < 0 | $BF_{-0}$ = 0.08 | [-0.15, 0.23] | *d* = 0.04 | *NA* |
| | All > 0 | $BF_{+0}$ = 0.16 | [-0.15, 0.23] | *d* = 0.04 | *NA* |
| Confidence pre | All ≠ 0 | $BF_{01}$ = 7.12 | [-0.12, 0.27] | *d* = 0.07 | *NA* |
| | Bias ≠ 0 | $BF_{10}$ = 1.17 | [0.00, 0.60] | *d* = 0.30 | *NA* |
| | No-bias ≠ 0 | $BF_{01}$ = 4.38 | [-0.38, 0.13] | *d* = -0.12 | [0.45, 1.50] |
| Sampling | All ≠ 0.377 | $BF_{01}$ = 5.51 | [1.19, 1.77] | *d* = 1.48 | [0.36, 1.50] |
| | All = 0.5 | $BF_{10}$ = 36.51 | [1.20, 1.78] | *d* = 1.48 | [0.02, 1.50] |
| Rel. contingency post | Bias < 0 | $BF_{-0}$ = 0.17 | [-0.30, 0.27] | *d* = -0.01 | *NA* |
| | Bias = 0 | $BF_{01}$ = 6.05 | [-0.30, 0.28] | *d* = -0.01 | [0.32, 1.50] |
| | No-bias < 0 | $BF_{-0}$ = 77,282 | [-1.03, -0.44] | *d* = -0.73 | [0.02, 1.50] |
| Cond. Estimates post | Bias < 0 | $BF_{-0}$ = 0.32 | [-0.39, 0.19] | *d* = -0.11 | *NA* |
| | Bias = 0 | $BF_{01}$ = 4.69 | [-0.39, 0.18] | *d* = -0.10 | [0.42, 1.50] |
| | No-bias < 0 | $BF_{-0}$ = 121.43 | [-0.75, -0.21] | *d* = -0.47 | [0.02, 1.50] |
| Confidence post | Bias ≠ 0 | $BF_{01}$ = 4.77 | [-0.39, 0.18] | *d* = -0.10 | [0.41, 1.50] |
| | No-bias ≠ 0 | $BF_{01}$ = 4.63 | [-0.37, 0.14] | *d* = -0.11 | [0.42, 1.50] |

A

# References

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology, 78*(1), 53-63. https://doi.org/10.1037/0022-3514.78.1.53

Abbott, K. R., & Sherratt, T. N. (2011). The evolution of superstition through optimal use of incomplete information. *Animal Behaviour, 82*(1), 85-92. https://doi.org/10.1016/j.anbehav.2011.04.002

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-38). Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*(3), 147-149. https://doi.org/10.3758/BF03334492

Alves, H., Koch, A., & Unkelbach, C. (2018). A Cognitive-Ecological Explanation of Intergroup Biases. *Psychological Science, 29*(7), 1126-1133. https://doi.org/10.1177/0956797618756862

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm [doi:10.1037/h0021966]. *Journal of Personality and Social Psychology, 2*(1), 1-9. https://doi.org/10.1037/h0021966

Asch, S. E. (1946). Forming impressions of personality [doi:10.1037/h0055756]. *The Journal of Abnormal and Social Psychology, 41*(3), 258-290. https://doi.org/10.1037/h0055756

Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., & Todorov, A. (2015). Thrill of victory or agony of defeat? Perceivers fail to utilize information in facial movements. *Emotion, 15*(6), 791-797. https://doi.org/10.1037/emo0000073

Bai, X., Fiske, S. T., & Griffiths, T. (2021). *Inaccurate Stereotypes from Rational Exploration*. PsyArXiv. https://doi.org/10.31234/osf.io/6hbvg

Bandura, A., & Kupers, C. J. (1964). Transmission of patterns of self-reinforcement through modeling. *The Journal of Abnormal and Social Psychology, 69*(1), 1-9. https://doi.org/10.1037/h0041187

Bargh, J. A. (1990). Goal and Intent: Goal-Directed Thought and Behavior Are Often Unintentional. *Psychological Inquiry, 1*(3), 248-251. https://doi.org/10.1207/s15327965pli0103_14

Barnes, P. M. M. A., Bloom, B., & Nahin, R. L. (2008). Complementary and alternative medicine use among adults and children; United States, 2007. U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention,

National Center for Health Statistics Retrieved from https://stacks.cdc.gov/view/cdc/5266

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beer, A.-M., Burlaka, I., Buskin, S., Kamenov, B., Pettenazzo, A., Popova, D., Riveros Huckstadt, M. P., Sakalinskas, V., & Oberbaum, M. (2016). Usage and Attitudes Towards Natural Remedies and Homeopathy in General Pediatrics: A Cross-Country Overview. *Global Pediatric Health, 3*, 2333794X15625409. https://doi.org/10.1177/2333794X15625409

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214-1221. https://doi.org/10.1038/nn1954

Bernoulli, J. (1713). *Ars conjectandi, opus posthumum* [The art of conjecturing, posthumous work]. Thurneysen Brothers.

Berridge, K. C. (2000). Reward learning: Reinforcement, incentives, and expectations. *Psychology of Learning and Motivation, 40*, 223-278. https://doi.org/http://dx.doi.org/10.1016/S0079-7421(00)80022-5

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist, 71*(8), 670-679. https://doi.org/10.1037/amp0000059

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology, 53*(3), 155-167. https://doi.org/10.1016/j.jmp.2008.05.006

Bohnlein, J., Altegoer, L., Muck, N. K., Roesmann, K., Redlich, R., Dannlowski, U., & Leehr, E. J. (2020). Factors influencing the success of exposure therapy for specific phobia: A systematic review. *Neuroscience & Biobehavioral Reviews, 108*, 796-820. https://doi.org/10.1016/j.neubiorev.2019.12.009

Bott, F. M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000840

Brunswik, E. (1952). *The conceptual framework of psychology. (Int. Encycl. unified Sci., v. 1, no. 10.)*. Univ. Chicago Press.

Burns, J. (2006). Racial stereotypes, stigma and trust in post-apartheid South Africa. *Economic Modelling, 23*(5), 805-821. https://doi.org/10.1016/j.econmod.2005.10.008

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 B: Biological Sciences, 362*(1481), 933-942. https://doi.org/10.1098/rstb.2007.2098

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences, 10*(7), 294-300. https://doi.org/10.1016/j.tics.2006.05.004

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: an inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10-23. https://doi.org/10.1016/j.brat.2014.04.006

Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science, 329*(5987), 47-50. https://doi.org/10.1126/science.1188595

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674. https://doi.org/10.1126/science.2648573

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*(1), 2-27. https://doi.org/10.1207/s15327957pspr0801_1

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin, 21*(10), 1083-1092. https://doi.org/10.1177/01461672952110009

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*(1), 152-164. https://doi.org/10.3758/BF03195749

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological Review, 112*(4), 951-978. https://doi.org/10.1037/0033-295X.112.4.951

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

Denrell, J., & Le Mens, G. (2011). Seeking positive experiences can produce illusory correlations. *Cognition, 119*(3), 313-324. https://doi.org/10.1016/j.cognition.2011.01.007

Denrell, J., & Le Mens, G. (2012). Social judgments from adaptive samples. In J. I. Krueger (Ed.), *Social judgment and decision making.* (pp. 151-169). Psychology Press.

Denrell, J., & March, J. G. (2001). Adaptation as Information Restriction: The Hot Stove Effect. *Organization Science, 12*(5), 523-538. https://doi.org/10.1287/orsc.12.5.523.10092

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*(1), 180-209. https://doi.org/10.1037/0033-295X.106.1.180

Ebbesen, E. B., Kjos, G. L., & Konečni, V. J. (1976). Spatial ecology: Its effects on the choice of friends and enemies. *Journal of Experimental Social Psychology, 12*(6), 505-518. https://doi.org/https://doi.org/10.1016/0022-1031(76)90030-5

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380–417. https://doi.org/https://doi.org/10.1037/h0053870

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*(5), 395-416. https://doi.org/10.1037/0033-295X.85.5.395

Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist simulation of attitude learning: asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin, 29*(10), 1221-1235. https://doi.org/10.1177/0146167203254605

Ernst, E. (2010). Homeopathy: what does the “best” evidence tell us? *Medical Journal of Australia, 192*(8), 458-460.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale, 56*(1), 58-63. https://doi.org/10.1037/h0087385

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*(2), 175-191. https://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*(3), 293-311. https://doi.org/10.1037/0022-3514.87.3.293

Feingold, A. (1992). Good-Looking People Are Not What We Think. *Psychological Bulletin, 111*(2), 304-341.

Fetchenhauer, D., & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science, 21*(2), 189-193. https://doi.org/10.1177/0956797609358586

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*(1), 193-214.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659-676. https://doi.org/10.1037/0033-295X.107.4.659

Fiedler, K. (2008). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 186-203. https://doi.org/10.1037/0278-7393.34.1.186

Fiedler, K. (2010). Pseudocontingencies can override genuine contingencies between multiple cues. *Psychonomic Bulletin & Review, 17*(4), 504-509. https://doi.org/10.3758/PBR.17.4.504

Fiedler, K. (2012). Meta-Cognitive Myopia and the Dilemmas of Inductive-Statistical Inference. In *The Psychology of Learning and Motivation* (pp. 1-55). https://doi.org/10.1016/b978-0-12-394293-7.00001-7

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology, 87*(4), 453-467. https://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*(1), 187-206. https://doi.org/10.1037/a0014480

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudocontingencies in a Simulated Classroom. *Journal of Personality and Social Psychology, 92*(4), 665-677.

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically Unwarranted but Smart Inferences. *Current Directions in Psychological Science, 22*(4), 324-329. https://doi.org/10.1177/0963721413480171

Fiedler, K., Walther, E., & Nickel, S. (1999). The auto-verification of social hypotheses: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology, 77*(1), 5-18. https://doi.org/10.1037/0022-3514.77.1.5

Fiedler, K., & Wänke, M. (2009). The cognitive-ecological approach to rationality in social psychology. *Social Cognition, 27*(5), 699-732. https://doi.org/10.1521/soco.2009.27.5.699

Fishbein, M., & Ajzen, I. (1975). *Intention and Behavior: An introduction to theory and research.* Addison-Wesley.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition, 109*(3), 416-422.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology, 68*, 101-128. https://doi.org/10.1146/annurev-psych-122414-033625

Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology, 58*(1), 129-149. https://doi.org/10.1111/bjso.12286

Gilboa, I., & Schmeidler, D. (1995). Case-Based Decision Theory. *The Quarterly Journal of Economics, 110*(3), 605-639.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In *Handbook of motivation and cognition: Foundations of social behavior, Vol. 2.* (pp. 53-92). The Guilford Press.

Gollwitzer, P. M., & Moskowitz, G. B. (1996). Goal Effects on Action and Cognition. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 361-399). Guilford.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science, 17*(9), 767-773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Harris, A. J., & Hahn, U. (2011). Unrealistic optimism about future life events: a cautionary note. *Psychological Review, 118*(1), 135-154. https://doi.org/10.1037/a0020997

Harris, C. (2020). *Repository for: Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments.* https://osf.io/j3ve4/

Harris, C., Aarts, H., Fiedler, K., & Custers, R. (2022a). Iterative Cycles in Psychology: How Hedonic Outcomes, Idiosyncratic Experiences, and Decisions by Experience Shape Our Subjective World. *Manuscript in preparation*.

Harris, C., Aarts, H., Fiedler, K., & Custers, R. (2022b). Missing Out by Pursuing Rewarding Outcomes: Why Initial Biases Can Lead to Persistent Suboptimal Choices. *Manuscript submitted for publication*.

Harris, C., & Custers, R. (2022). Biased Preferences Through Exploitation. In K. Fiedler, P. Juslin, & J. Denrell (Eds.), *Sampling in Judgement and Decision Making*. Invited chapter in press.

Harris, C., Fiedler, K., & Custers, R. (2022). When Biases Beget Biases: Erroneous First Impressions of Choice Options Thrive in Settings With Graded Outcomes. *Manuscript submitted for publication*.

Harris, C., Fiedler, K., Marien, H., & Custers, R. (2020). Biased preferences through exploitation: How initial biases are consolidated in reward-rich environments. *Journal of Experimental Psychology: General, 149*(10), 1855–1877.

Harris, C., Prager, J., Hess, P., & Fiedler, K. (2021). When Small Samples or Large Samples Carry Stronger Information: Evidence from the Virtual Classroom Paradigm. *Manuscript in preparation*.

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

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

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280-1300. https://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*(1), 46-54. https://doi.org/https://doi.org/10.1016/j.tics.2014.10.004

Hinson, J. M., & Staddon, J. E. R. (1983). Matching, Maximizing, and Hill-Climbing. *Journal of the Experimental Analysis of Behavior, 40*(3), 321-331. https://doi.org/10.1901/jeab.1983.40-321

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The Two Settings of Kind and Wicked Learning Environments. *Current Directions in Psychological Science, 24*(5), 379-385. https://doi.org/10.1177/0963721415591878

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2018). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*. https://doi.org/10.1037/met0000201

Hotaling, J. M., Navarro, D. J., & Newell, B. R. (2020). Skilled bandits: Learning to choose in a reactive world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000981

Jaeger, B., Harris, C., Custers, R., & Todorov, A. (2021). Biased partner selection propagates stereotypes and inequalities in social exchange. *Manuscript submitted for publication*.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*(1), 1-17. https://doi.org/10.1037/h0093874

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*(2), 250 - 274.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. https://doi.org/10.2307/1914185

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*(3), 263-269. https://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*(2), 397-402. https://doi.org/10.1037/0033-295X.107.2.397

Kareev, Y., Avrahami, J., & Fiedler, K. (2014). Strategic interactions, affective reactions, and fast adaptations. *Journal of Experimental Psychology: General, 143*(3), 1112-1126. https://doi.org/10.1037/a0034641

Kareev, Y., Fiedler, K., & Avrahami, J. (2009). Base rates, contingencies, and prediction behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 371-380. https://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*(3), 278-287. https://doi.org/10.1037/0096-3445.126.3.278

Kasper, J., Fiedler, K., Kutzner, F., & Harris, C. (2021). On the Role of Exploitation and Exploration Strategies in the Maintenance of Cognitive Biases: Beyond Hedonic Motives. *Manuscript in preparation*.

Kauff, M., Beneda, M., Paolini, S., Bilewicz, M., Kotzur, P., O'Donnell, A. W., Stevenson, C., Wagner, U., & Christ, O. (2020). How do we get people into contact? Predictors of intergroup contact and drivers of contact seeking. *Journal of Social Issues, 77*(1), 38-63. https://doi.org/10.1111/josi.12398

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). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60315-1

Klayman, J., & Ha, Y.-w. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211-228. https://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*(2), 145-168. https://doi.org/10.1016/0022-1031(92)90036-J

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244. https://doi.org/10.1038/nature02169

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal-systems. *Advances in Experimental Social Psychology*, 311-378.

Kruschke, J. K. (2003). Attention in Learning. *Current Directions in Psychological Science, 12*(5), 171-175. https://doi.org/10.1111/1467-8721.01254

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. https://doi.org/10.1037/0033-2909.108.3.480

Kutzner, F. (2009). *Pseudocontingencies - rule based and associative* University of Heidelberg]. Heidelberg, Germany. http://archiv.ub.uni-heidelberg.de/volltextserver/9912/1/Dissertation_Kutzner_UB.pdf

Kutzner, F., & Fiedler, K. (2017). Stereotypes as Pseudocontingencies. *European Review of Social Psychology, 28*(1), 1-49. https://doi.org/10.1080/10463283.2016.1260238

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lave, C. A., & March, J. G. (1993). *An Introduction to Models in the Social Sciences*. UPA. https://books.google.nl/books?id=LfyBAAAAQBAJ

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the 'naivety' assumption in sampling explanations of judgment biases. *Psychological Review, 118*(2), 379-392. https://doi.org/10.1037/a0023010

Leiner, D. J. (2020). *SoSci Survey* (Version 3.2.05-i) [Computer software]. https://www.soscisurvey.de

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*(1), 129-137. https://doi.org/10.1037/h0035519

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*(11), 2098-2109. https://doi.org/10.1037/0022-3514.37.11.2098

Marien, H., Custers, R., & Aarts, H. (2019). Studying Human Habits in Societal Context: Examining Support for a Basic Stimulus–Response Mechanism. *Current Directions in Psychological Science, 28*(6), 614-618. https://doi.org/10.1177/0963721419868211

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

Meiser, T., & Hewstone, M. (2004). Cognitive processes in stereotype formation: the role of correct contingency learning for biased group judgments. *Journal of Personality and Social Psychology, 87*(5), 599-614. https://doi.org/10.1037/0022-3514.87.5.599

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*(3), 315-336. https://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*(1), 50 - 67. https://doi.org/10.1037/xlm0000432

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science, 8*(6), 423-429. https://doi.org/10.1111/j.1467-9280.1997.tb00455.x

Mischel, W. (1974). Processes in Delay of Gratification. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 7, pp. 249-292). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2601(08)60039-8

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]. https://CRAN.R-project.org/package=BayesFactor

Navarro, D. J., Newell, B. R., & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore-exploit dilemma in static and dynamic environments. *Cognitive Psychology, 85*, 43-77. https://doi.org/10.1016/j.cogpsych.2016.01.001

Newell, B. R., Lagnado, D. A., & Shank, D. R. (2015). *Straight Choices: The Psychology of Decision Making* (2nd ed.). Psychology Press. https://doi.org/https://doi-org.proxy.library.uu.nl/10.4324/9781315727080

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175 - 220.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience, 14*(9), 1105-1107. https://doi.org/10.1038/nn.2886

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature, 364*(6432), 56-58. https://doi.org/10.1038/364056a0

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608-631. https://doi.org/10.1037/0033-295X.101.4.608

Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: evidence from a secondary task. *Cognition, 118*(2), 274-279. https://doi.org/10.1016/j.cognition.2010.11.009

Paluck, E. L., Green, S. A., & Green, D. P. (2018). The contact hypothesis re-evaluated. *Behavioural Public Policy, 3*(02), 129-158. https://doi.org/10.1017/bpp.2018.25

Paolini, S., White, F. A., Tropp, L. R., Turner, R. N., Page-Gould, E., Barlow, F. K., & Gómez, Á. (2021). Intergroup contact research in the 21st century: Lessons learned and forward progress if we remain open. *Journal of Social Issues, 77*(1), 11-37. https://doi.org/10.1111/josi.12427

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *J Pers Soc Psychol, 90*(5), 751-783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922-934. https://doi.org/10.1002/ejsp.504

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*, 46-63. https://doi.org/https://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. https://doi.org/https://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*(3), 379-397. https://doi.org/10.1037/pspa0000123

Preciado, P., Snijders, T. A., Burk, W. J., Stattin, H., & Kerr, M. (2012). Does Proximity Matter? Distance Dependence of Adolescent Friendships. *Social Networks, 34*(1), 18-31. https://doi.org/10.1016/j.socnet.2011.01.002

R Core Team. (2018). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Relton, C., Cooper, K., Viksveen, P., Fibert, P., & Thomas, K. (2017). Prevalence of homeopathy use by the general population worldwide: a systematic review. *Homeopathy, 106*(2), 69-78. https://doi.org/https://doi.org/10.1016/j.homp.2017.03.002

Rescorla, R., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non reinforcement. In A. H. Black

& W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64 - 99). Appleton-Century-Crofts.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18*(1), 89-107. https://doi.org/10.1016/j.lindif.2007.03.011

Savage, L. J. (1954). *The Foundations of Statistics*. NY Wiley.

Shang, A., Huwiler-Müntener, K., Nartey, L., Jüni, P., Dörig, S., Sterne, J. A. C., Pewsner, D., & Egger, M. (2005). Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *The Lancet, 366*(9487), 726-732. https://doi.org/https://doi.org/10.1016/S0140-6736(05)67177-2

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience, 14*(11), 1475-1479. https://doi.org/10.1038/nn.2949

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour, 4*(1), 14-19. https://doi.org/10.1038/s41562-019-0793-1

Sheeran, P., & Webb, T. L. (2016). The Intention–Behavior Gap. *Social and Personality Psychology Compass, 10*(9), 503-518. https://doi.org/https://doi.org/10.1111/spc3.12265

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. https://doi.org/http://dx.doi.org/10.2139/ssrn.2160588

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118. https://doi.org/10.2307/1884852

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138. https://doi.org/10.1037/h0042769

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology, 38*(2), 168-172.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science, 7*(2), 351-367. https://doi.org/10.1111/tops.12145

Staudinger, M. R., & Büchel, C. (2013). How initial confirmatory experience potentiates the detrimental influence of bad advice. *Neuroimage, 76*, 125-133. https://doi.org/10.1016/j.neuroimage.2013.02.074

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology, 53*(3), 168-179. https://doi.org/10.1016/j.jmp.2008.11.002

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology, 39*, 212-222. https://doi.org/10.2307/1415413

Tolman, E. C. (1934). Theories of learning. In *Comparative psychology.* (pp. 367-408). Prentice-Hall, Inc. https://doi.org/10.1037/11453-012

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110. https://doi.org/10.1037/h0031322

Verplanken, B. (Ed.). (2018). *The Psychology of Habit* (1 ed.). Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-319-97529-0.

Vogel, T., & Kutzner, F. (2017). Pseudocontingencies in Consumer Choice: Preference for Prevalent Product Categories Decreases with Decreasing Set Quality. *Journal of Behavioral Decision Making, 30*(5), 1193-1205. https://doi.org/doi:10.1002/bdm.2034

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd rev. ed.).

Vulkan, N. (2000). An Economist's Perspective on Probability Matching. *Journal of Economic Surveys, 14*, 101-118. https://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]. https://CRAN.R-project.org/package=dplyr

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology: General, 143*(6), 2074-2081. https://doi.org/10.1037/a0038199

Wilson, R. K., & Eckel, C. C. (2016). Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly, 59*(2), 189-202. https://doi.org/10.1177/106591290605900202

Winet, Y. K., Tu, Y., Choshen-Hillel, S., & Fishbach, A. (2020). Social exploration: When people deviate from options explored by others. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000350

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology, 67*(1), 289-314. https://doi.org/10.1146/annurev-psych-122414-033417

Yu, A. J. (2007). Adaptive Behavior: Humans Act as Bayesian Learners. *Current Biology, 17*(22), R977-R980. https://doi.org/10.1016/j.cub.2007.09.007

# Summary in Dutch (Nederlandse samenvatting)

Stel je voor, je gaat op een eerste date. Als de date goed gaat, is er een kans dat er vervolgdates komen die potentieel in een relatie kunnen eindigen. Maar als de eerste date niet goed gaat, daalt de kans voor een vervolgdate enorm. Misschien was dit de juiste keuze, maar misschien was er wel een kans voor een gedeelde toekomst geweest. In dit geval zorgt een negatieve ervaring op de eerste date ervoor dat je jezelf nooit de kans geeft om meer over de andere persoon te leren en je eerste indruk bij te stellen.

Dit voorbeeld, gebaseerd op Denrell (2005), maakt duidelijk hoe ons gedrag en de keuzes die we maken een invloed hebben op wat voor informatie en feedback we tegen zouden kunnen komen terwijl we met de wereld om ons heen interacteren. Als we op vervolgdates zouden gaan, zouden we meer over de andere persoon leren. Wellicht blijkt dan de persoon wel bij je te passen. Door niet op vervolgdates te gaan ontnemen we ons de mogelijkheid om onze indruk van deze persoon bij te stellen. Deze eenvoudige relatie tussen ons gedrag, informatie die we tegenkomen, en hoe we vervolgens onze overtuigingen kunnen bijstellen, vormt de kern van deze dissertatie. Vaak leren mensen door de positieve of negatieve uitkomsten van de keuzes die ze maken: Was de date goed of slecht? Dat zegt iets over de andere persoon en de connectie die we hadden en dus over in hoeverre we matchen. Is deze persoon betrouwbaar?

Maar dit betekent dat iedere keer dat we herhaaldelijke keuzes maken, we een fundamentele afweging moeten maken: Willen we meer over alternatieven leren of willen we voor de beste uitkomst gaan? Deze afweging wordt ook de “exploration-exploitation-tradeoff” genoemd. Een mogelijkheid is, om tussen verschillende opties te variëren. Het voordeel van de strategie van het exploreren van alternatieven is dat je meer informatie over de keuzeopties leert of misschien zelfs een nieuwe, betere ontdekt. Je krijgt dus meer informatie over alle opties, maar dat zal tegelijkertijd betekenen dat je tussendoor ook minder goede opties kiest. De andere mogelijkheid is om iedere keer de vermeende beste optie te kiezen en deze dus te exploiteren. Het voordeel van deze strategie is dat als dit inderdaad de beste optie is, je de grootste kans op een positieve uitkomst hebt. Maar dit betekent ook dat je niets leert over de alternatieven en je dus niet merkt als er iets verandert of als een ander alternatief zelfs nog beter zou zijn. Exploratie is dus vaak beter om te leren, terwijl exploitatie vaak beter is om maximaal veel positieve uitkomsten te krijgen.

Zowel onderzoek als ons evolutionair succes laat ruimschoots zien dat mensen meestal best goed zijn in het afwegen van deze twee strategieën en veelal voor een slimme combinatie van deze twee strategieën kiezen. Maar tegelijkertijd zitten ook een flink aantal haken en ogen aan het bestaande onderzoek naar keuzegedrag: Een centraal probleem is dat onderzoek naar keuzegedrag vaak informatie aan proefpersonen voedt en het keuzegedrag van proefpersonen verder geen invloed heeft op de informatie die ze daaropvolgend te zien krijgen. Echter, dit negeert de wisselwerking tussen de keuzes die proefpersonen maken en wat voor informatie ze überhaupt tegen zouden kunnen komen. Hierdoor vervalt het dilemma om constant een keuze te maken met betrekking tot de exploration-exploitation-tradeoff compleet, terwijl het in ons daadwerkelijke keuzegedrag een centrale rol speelt. Als gevolg kan dit onderzoek vaak niet goed verklaren waarom mensen soms in dingen blijven geloven die niet kloppen, zelfs als proefpersonen herhaaldelijk de mogelijkheid hebben om hun overtuigingen te testen. Veel onderzoek biedt verklaringen over hoe bijvoorbeeld foutieve eerste indrukken, stereotypes of bijgeloof kunnen ontstaan. Maar waarom blijven deze bestaan ondanks dat we onze overtuigingen herhaaldelijk zouden kunnen testen en aanpassen?

In mijn proefschrift stel ik dat dit juist samenhangt met de exploration-exploitation-tradeoff. Mensen leren door de ervaringen die ze maken, maar deze ervaringen hangen direct af van de keuzes die ze bij voorbaat maken. Door het actieve keuzegedrag bepalen mensen wat voor informatie ze mogelijk tegen zouden kunnen komen. Als je in een situatie bent waar de uitkomsten van je keuzegedrag voornamelijk positief zijn en dus de meeste uitkomsten voordelig zijn, dan kan het heel snel lijken alsof je al de beste strategie hebt gevonden. Uiteraard, door verder te exploreren zou je kunnen ervaren dat andere opties mogelijk even goed of zelfs beter zijn, maar je komt al voornamelijk positieve uitkomsten tegen – waarom dus verder zoeken? Als ons doel zou zijn om met eindeloze resources (denk aan geld, tijd, etc.) over de keuzealternatieven te leren, dan zou je zeker verder zoeken. Maar ik stel dat dit vaak niet ons doel is. In plaats daarvan zijn we op zoek naar positieve uitkomsten. We gaan niet op dates om over alle alternatieven te leren, maar om een leuke avond te hebben met een date. En we zouden bijvoorbeeld ook niet zo snel naar een vies lijkende broodjeszaak gaan puur om iets over deze broodjeszaak te leren, maar zijn eerder geïnteresseerd in waar we waarschijnlijk wel een lekker broodje krijgen. En als dus een bepaalde optie de betere optie lijkt te zijn omdat we als gevolg vaak positieve uitkomsten ervaren, dan hebben we de neiging om bij deze optie te blijven. De consequentie is dan echter wel dat we niet (genoeg) over alternatieven leren en dus ook niet onze (foutieve) eerste indruk kunnen overkomen.

In de hoofdstukken van mijn proefschrift presenteer ik mijn empirisch onderzoek naar deze interactie tussen overtuigingen en keuzegedrag en beschrijf ik hoe mijn werk in de grotere context van het vakgebied kan worden geplaatst. Ik ga in op de vraag of en hoe deze wisselwerking tussen gedrag, bewijsmateriaal en overtuigingen kan leiden tot hardnekkige vooroordelen, ondanks de vele mogelijkheden om iemands overtuigingen op de proef te stellen. Ik zal eindigen met het argument dat mensen uiteindelijk hun eigen idiosyncratische ervaring van de wereld vormgeven. Na een inleiding in Hoofdstuk 1 laat ik in Hoofdstuk 2-5 mijn empirische werk zien. Hiervoor gebruik ik variaties van een two-armed bandit task, wat betekent dat proefpersonen herhaaldelijk tussen twee opties moeten kiezen om punten te verdienen, wat later verrekend wordt met hoeveel ze worden uitbetaald. Ik zorg ervoor dat proefpersonen in het begin denken dat een bepaalde optie beter is dan de andere, terwijl ze feitelijk even goed zijn. Hoofdstuk 2 staat hier centraal omdat ik hier de te gronde liggende theorie in detail uitleg en de invloed laat zien die de omgeving heeft op het wel of niet overeind houden van een initiële bias: Als de uitkomsten voornamelijk positief zijn, blijft een initiële bias bestaan. Echter, als de uitkomsten voornamelijk negatief zijn, lukt het de proefpersonen om de initiële bias te overkomen. In Hoofstuk 3-5 laat ik variaties van deze eerste experimenten zien: Zelfs als uitkomsten niet binair zijn (winst vs. verlies) maar gradueel, kan hetzelfde patroon ontstaan. Dan blijkt zelfs dat als de alternatieve optie beter is, proefpersonen bij de in hun ogen betere optie blijven hangen. Tenslotte rond ik mijn proefschrift af met Hoofdstuk 6. Hier schets ik in een theoretisch paper het grotere plaatje waarin ik mijn proefschrift zie. Ik stel een iteratieve cyclus voor waarin het kiezen van opties, het bijstellen van je mentale representatie en het maken van de daaropvolgende keuze van elkaar afhangen en elkaar beïnvloeden. Deze onderlinge afhankelijkheid kan verklaren waarom we soms aan foutieve gedachtes vasthouden of waarom stereotypes en bijgeloof blijven bestaan. In mijn proefschrift stel ik dat door het actief maken van keuzes, we uiteindelijk allemaal een heel persoonlijke belevenis van de wereld creëren.

# Acknowledgements

# Curriculum Vitae

Chris Harris was born in Malsch, Germany on the 16th of January, 1990. He obtained his bachelor's degree in psychology at the University of Heidelberg in 2013. Following a one-year visit at the University of Connecticut, he obtained his master's degree in psychology at the University of Heidelberg in 2016. That same year, he began to work on his PhD project that is described in this dissertation under the supervision of prof. Klaus Fiedler from the University of Heidelberg and prof. Henk Aarts and dr. Ruud Custers from Utrecht University. In 2018, he obtained a grant by the German Academic Exchange Service (DAAD) which allowed him to continue his work at Utrecht University in the Netherlands. From 2019 until 2021 he worked as a lecturer in psychology at Utrecht University. He is currently working as assistant professor at the department for Social, Health, and Organizational Psychology at Utrecht University.