I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13821

Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events

Christoph K. Becker
Tigran Melkonyan
Eugenio Proto
Andis Sofianos
Stefan T. Trautmann

OCTOBER 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13821

Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events

Christoph K. Becker
University of Heidelberg

Tigran Melkonyan
University of Alabama, Tuscaloosa

Eugenio Proto
University of Glasgow, CEPR, IZA and CesIfo

Andis Sofianos
University of Heidelberg

Stefan T. Trautmann
University of Heidelberg

OCTOBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events*

Bayesian Updating is the dominant theory of learning in economics. The theory is silent about how individuals react to events that were previously unforeseeable or unforeseen. Recent theoretical literature has put forth axiomatic frameworks to analyze the unknown. In particular, we test if subjects update their beliefs in a way that is consistent "reverse Bayesian", which ensures that the old information is used correctly after an unforeseen event materializes. We find that participants do not systematically deviate from reverse Bayesianism, but they do not seem to expect an unknown event when this is reasonably unforeseeable, in two pre-registered experiments that entail unforeseen events. We argue that participants deviate less from the reverse Bayesian updating than from the usual Bayesian updating. We provide further evidence on the moderators of belief updating.

JEL Classification: C11, C91, D83, D84

Keywords: reverse Bayesianism, unforeseen, unawareness, Bayesian Updating

Corresponding author:
Eugenio Proto
Adam Smith Business School
University of Glasgow
University Avenue
Glasgow, G12 8QQ
United Kingdom

E-mail: eugenio.proto@gmail.com

* The authors thank several colleagues for discussions and insights on this project. In particular, Pasquale Cirillo, Jürgen Eichberger, Peter Hammond, Andrea Isoni, Edi Karni, Joerg Oechssler, Daniel Sgroi. We also thank participants of the ESA 2020 Global for helpful comments. We are very grateful to J. Philipp Reiss for the use of the Karlsruhe Decision & Design Lab.

1 Introduction

We live in a world where scientific progress, human activities, and events outside of our control constantly lead to discoveries and observations of *unforeseen* phenomena that fundamentally change our worldviews and behavior. Examples of such phenomena include global pandemics, political and economic crises as well as scientific groundbreaking discoveries – of which more specific examples we offer below.

These scenarios involve zero probability events where it is well-known that the Bayes rule is silent about decision-makers' belief updating and subsequent reactions; a number of theories have been advanced to explain behavior under such circumstances.[1] The focus of our study is an experimental investigation of how individuals act in the face of unforeseeable and unforeseen events and react after these events materialize. To which degree are they aware of the possibility of unforeseen events?

Karni and Vierø (2013, 2015, 2017) and Karni et al. (2020) propose a framework that axiomatizes and analyzes growing awareness of unforeseeable and unforeseen events. Their central normative postulate is *reverse Bayesianism* (Karni and Vierø, 2013) according to which decision-makers should react to probability-zero events by proportionately shifting probability mass from the prior non-null events to the prior null but posterior non-null events. That is, a reverse Bayesian's construction of a new universe maintains consistency with the old structure.

In many environments, reverse Bayesianism imposes a rationality constraint on the process of updating beliefs following null events. Suppose, for example, a decision-maker faces an urn with 50 black marbles and 50 white marbles and this information can be objectively verified. The decision-maker makes bets on the color of a randomly drawn marble. Consider the scenario where the decision-maker witnesses some number (known or unknown) of red marbles being unexpectedly added to the urn. Under this design, the contents of the original urn with two colors are part of the updated urn with three colors. Thus, a part of the "old world" remains a part of the "new world". Hence, irrespective of the number of red marbles added to the urn, a rational updating rule (that does not dismiss or improperly use any objective information) requires that the decision-maker's posterior beliefs put equal probability weight on white and black marbles, which is the (only) requirement of reverse Bayesianism in this example.

The present paper develops two experiments to determine whether individuals adhere to reverse Bayesianism and, more generally, to determine formation of beliefs under growing awareness. We design our experiments so that the new and unexpected environment retains some parts of the old world which also allows us to address rationality of belief updating. In addition, we test how belief formation and updating is moderated by the environment of the decision situation. According to our knowledge, there are no studies of this nature in the received literature. We find that behavior in both experiments is consistent with reverse Bayesianism, despite the fact that the participants

[1] Dekel et al. (1998); Modica and Rustichini (1999); Heifetz et al. (2006); Halpern and Rêgo (2008); Kochov (2010); Ortoleva (2012); Grant and Quiggin (2013); Karni and Vierø (2013); Schipper (2013); Dominiac and Tserenjigmid (2017); Grant et al. (2017); Dietrich (2018)

exhibited some commonly observed judgmental biases. Based on our findings, reverse Bayesianism seems to be a natural updating rule for decision-makers.

Situations with unforeseen events are abundant and are characterized by being more or less objectively unforeseeable and entail the various "unknown unknowns", a term famously coined by the former US Secretary of Defense Donald Rumsfeld. Even when we can imagine rough outlines of a phenomenon or even when we have a rather precise understanding of its characteristics, we often overlook it in the construction of our universe or include it in the description of the universe but render it as impossible. That is, some aspects of the choice environment may be unforeseen.

Examples of objectively unforeseeable and unforeseen events are ubiquitous and include scientific breakthroughs, technological discoveries, and various global events. It is hard to disagree that the concept of antimatter was unforeseeable before the first particulate theories of matter were proposed by ancient philosophers Kanada (around 6^{th} century BCE) Leucippus (around 490 BCE) and Democritus (around 470-380 BCE) and before the word and notion "energy" first appeared in the works of Aristotle in the 4^{th} century BCE. Even if it was foreseeable to selected few after these scientific breakthroughs, the possibility of the opposite of normal matter was unforeseen by almost all until William Hicks and Karl Pearson proposed it between 1880 and 1890.

In some cases, the distinction between unforeseeable and unforeseen is blurred and individual-specific. As a result of differences in knowledge and cognitive capacity, what is foreseeable for some people might be unforeseeable to others. A very notable example of an event that was foreseeable but unforeseen by many is COVID-19. Numerous scientists and observers including Bill Gates have repeatedly warned us about the possibility of a disastrous pandemic. However, these warnings have been largely ignored by many policy makers, public health officials, and economic decision-makers.

The financial crisis of 2007 had a similar nature. It was foreseeable but unforeseen by most. Lehman Brothers assumed in 2005 that the worst-case scenario in the housing market was a temporary price depreciation of 5% over the next three years, followed by a rebound and price increases of 5% thereafter (Gennaioli and Shleifer, 2018, p. 52). We all know their fate and what happened to many other financial institutions that discarded the possibility of catastrophic financial and economic events. The fall of the Berlin Wall was also undoubtedly foreseeable. However, it was unforeseen by many political leaders and individuals. Witnessed time and time again, unforeseeable and unforeseen events have large consequences and lead to seismic behavioral changes. Failure to recognize the possibility of such events, no matter how unlikely they may seem, can impair the ability to mitigate adverse events and to take advantage of new opportunities.[2]

All of these examples entail, to a varying degree, *unawareness* on the part of decision-makers about various aspects of the choice environment. In some cases, this unawareness may arise because some states of nature are simply unforeseeable. In others, they are foreseeable but decision-makers fail to incorporate them in their model of the universe, rendering these states unforeseen.

In the controlled environment of an experiment it is virtually impossible to generate objectively

[2]These scenarios should also be distinguished from those unforeseens where a decision-maker contemplates a state of nature but assigns zero probability to that state. In contrast to the first two scenarios, this latter case corresponds to an aware decision-maker.

unforeseeable events. Our experimental designs involve events that are more or less unforeseeable for most of the participants. We find that participants generally do not expect the unknown when this is reasonably unforeseeable, but find evidence that some expect an unknown event when this is reasonably foreseeable. In our participant sample, cognitive abilities do not seem to generally affect formation and updating of beliefs.

In the first experiment, we analyze how participants face a reasonably unforeseeable event. Participants observe sufficiently many draws from an urn in the beginning of the experiment to form a good understanding of the prizes that are contained. In the course of the experiment, we elicit the beliefs about the content of the urn as well as willingness to sell a gamble that pays according to a prize randomly drawn from the urn. The experiment has two treatments. In both treatments, the reasonably unforeseeable event is the introduction a new urn (with new prize(s)) previously hidden and the subsequent addition of its content to the original urn. In one of the treatments, the participants are forewarned about the possibility that new prizes may be added while in the other treatment they are not.[3] Thus, there is room for the participants to be aware of such a possibility in the former treatment while reasonably not in the latter treatment. In both treatments, the emergence of a new urn may be unforeseen. We find that prior to encountering this surprise in the form of a new urn the participants in both treatments on average place a zero probability on all prizes other those they have previously observed. However, after witnessing the surprise, the event that something that has not yet been observed may materialize becomes non-null. In addition to this awareness of the unknown result and our test for reverse Bayesianism, we investigate how the nature of the surprise (large versus small prizes in the new urn) affects posterior beliefs. That is, we explore whether the surprise instills hope or fear. We find that after encountering a new urn participants place a higher valuation on the gamble that pays according to a random draw from the updated urn and that this holds even after they observe a negative surprise, although the valuation is lower in this case than after a positive surprise.

In the second experiment, the new events are reasonably foreseeable. Participants explore an urn by sequentially extracting marbles from it. They receive no information on which colors are possible, so that the new information is represented by so far unobserved colors. Participants in the second experiment are not systematically biased and we again find that beliefs are updated in accordance with reverse Bayesianism. Importantly, a large proportion of the participants appear to reason as perfect reverse Bayesians and we argue that participants deviate less from the reverse Bayesian updating than from the usual Bayesian updating. Additionally, we find that participants, irrespective of the number of unknown events experienced, as they sample more they lower their perceived likelihood of further unknown events. Nevertheless, about one third of the participants still expect a yet unobserved color even after 30 draws.

The rest of the paper is organized as follows. In section 2 we provide a theoretical framework and derive the main hypotheses that will be tested in both experiments. In section 3 we present the design and results of experiment 1. Section 4 is dedicated to experiment 2, design and results.

[3] As it will be clearer below, a main reason of this differentiation is to avoid any deception of our participants.

Section 5 provides a general discussion of the results and concludes the paper. In the appendix we include the experimental instructions and some additional analysis for each experiment.

2 Theoretical Background and Hypotheses

Theories of growing awareness about unforeseen events usually belong to one of three branches of literature: the epistemic approach, the game-theoretic or interactive decision making approach, or the choice-theoretic approach (Ortoleva, 2012; Karni and Vierø, 2013). In this paper we follow the approach first presented in Karni and Vierø (2013), which belongs to the third category. Karni and Vierø (2015) extends this approach from Expected Utility Theory to Probabilistic Sophistication (Machina and Schmeidler, 1992), which separates preferences from probabilistic beliefs. Karni and Vierø (2017) further generalizes the approach, while Karni et al. (2020) supplements the model by allowing states to become null after new discoveries.

Following Karni and Vierø (2017), let A denote a finite, non-empty set of actions and C_0 denote a finite, non-empty set of *feasible consequences*. It is instructive to use the example presented in Chambers and Hayashi (2018) to illustrate Karni and Vierø (2017)'s framework; accordingly A can represent the schools, a and b, an hypothetical student who would like to pursue a PhD in Economics can attend. While C_0 represents the consequences of her choice in terms of either getting a job at a research university, R, or working in a teaching college, T, after attending either school a or b. Let also $x = \neg C_0$ denote an abstract residual consequence, which stands for "something other than what the decision-maker can describe" – in our example a third type of institution our PhD student could find a job at, but she is not aware of.

Then, $\hat{C}_0 = C_0 \cup \{x\}$ and A together define the set of *augmented conceivable state space* via $\hat{C_0^A} := \{s : A \to \hat{C}_0\}$. That is, the *augmented conceivable state space* takes into account the possibility that an action may lead to the "everything else" consequence x. Karni and Vierø (2017) also define the *fully describable conceivable states* as $C_0^A := \{s : A \to C_0\}$, where the mapping's image is restricted only to describable consequences in the set of feasible consequences C_0.

The augmented conceivable state space can be expanded by observing a new consequence $c' \notin C_0$. Following Chambers and Hayashi (2018)'s example, the student can subsequently realize that, as a third consequence, she might end up with a job, P, in the private sector. The set of feasible consequences then expands to $C_1 = C_0 \cup \{c'\}$. Furthermore, C_1^A and $\hat{C_1^A}$ can be defined analogously to C_0^A and $\hat{C_0^A}$, respectively.

Denote π_0 and π_1 as probability measures defined on C_0^A and C_1^A, respectively before and after a new consequence is observed. Together with the axioms presented in Machina and Schmeidler (1995), which ensure probabilistic sophistication, their axiom for invariant risk preferences in Karni and Vierø (2013), and two additional axioms on refinement consistency, the framework yields the following result for all $s, s' \in C_0^A$:

$$\frac{\pi_0(s)}{\pi_0(s')} = \frac{\pi_1(s)}{\pi_1(s')},$$

That is, the model in Karni and Vierø (2013, 2015, 2017); Karni et al. (2020), implies that it is optimal for the decision maker to hold the ratio of probability estimates for known outcomes constant after observing an unforeseen outcome. Under classical Bayesian updating, new information shrinks the state space, for example, by excluding outcomes that had previously been assigned a positive prior probability (Karni and Vierø, 2013). In contrast, this model focuses on the reverse situation, where new information can expand the state space. Hence, the name "reverse Bayesianism."

Once a new consequence is discovered, this also affects the belief about further possible outcomes now captured by x. Observing a new outcome can have either an increasing or decreasing effect (or none) on the decision maker's awareness about unforeseen events: On the one hand, it is possible that the discovery of new consequences decreases the amount of remaining unforeseen consequences. On the other hand, discovering a new consequence could highlight that there are still unforeseen consequences to uncover. Accordingly, the model allows for both, either a decrease or increase in the probability assigned to the residual.

Based on this framework, we will test if the normative reverse Bayesian model presented in Karni and Vierø (2013, 2015, 2017); Karni et al. (2020) also matches the actual behavior of participants in an incentivized decision-making experiment. In both the experiments we present in the next sections, we simplify the framework presented in the current section by not allowing the choice of the urn or its composition to the participants, both of which are chosen by the experimenter. They are instead allowed to state their beliefs about the likelihood of the different consequences (respectively prizes and colors of the marbles in experiment 1 and experiment 2) and express their willingness to accept (WTA) for the urns using a standard truth-telling incentivizing procedure.

For experiment 1, the probability of an unknown prize is elicited implicitly by allowing participants to state probability estimates of the other known prizes not summing up to 1. The set of feasible consequences is given by different prizes in an urn. For experiment 2, the probability of an unknown event is elicited explicitly by asking participants to state their belief about "any other possible outcome". The set of feasible consequences is given by differently colored marbles in an urn. We differentiate between the *original urn* (before a new outcome is observed) and the *updated urn* (after a new outcome is observed).

In the following we denote the probability estimates of each participant for a given outcome i with $\hat{p}_i^o$ and $\hat{p}_i^u$, for the original and the updated urn respectively. The residual estimate is denoted by $\hat{p}_x^o$ and $\hat{p}_x^u$. Our two experiments will test the following main hypothesis:

Hypothesis 1. *Participants update their beliefs according to reverse Bayesianism. That is, for any $\hat{p}_i^o, \hat{p}_i^u$ and any outcomes $i, i' \in C_0^F$:*

$$\frac{\hat{p}_i^o}{\hat{p}_{i'}^o} = \frac{\hat{p}_i^u}{\hat{p}_{i'}^u}$$

In the analysis of the two experiments, we will refer to the difference between the ratios before and after the update obtained from the various tests as ΔR:

$$\Delta R = \frac{\hat{p}^u_i}{\hat{p}^u_{i'}} - \frac{\hat{p}^o_i}{\hat{p}^o_{i'}}$$

.

In some of our experimental treatments, we explicitly rule out the possibility of unforeseen consequences and inform the participants about this. Thus, if the participants believe this, then $\{x\}$, should be empty and $\hat{C}_0 = C_0$. Specifically, we will test the following:

Hypothesis 2. *(i) The residual estimate* $\hat{p}^j_x = 0$ *in treatments where unforeseen consequences are ruled out so that* $\{x\}$ *is empty and* $\hat{C}_0 = C_0$. *(ii) The restidual estimate* $\hat{p}^i_x > 0$ *in treatments where unforeseen consequences are not ruled out.*

Furthermore, as outlined above, the model presented in Karni and Vierø (2017) allows for decision makers to increase their residual in either direction after observing a new, unforeseen event. Accordingly, we test the following additional hypothesis.

Hypothesis 3. *Participants will not adjust their residual belief after an unforeseen event:*

$$\hat{p}^u_x - \hat{p}^o_x = 0$$

3 Experiment 1

3.1 Design

Experiment 1 elicits beliefs and valuations of prospects before and after encountering a reasonably unforeseeable event (we will omit the word "reasonably" henceforth). We test reverse Bayesianism in two conditions, *Information surprise (IS)* and *Payment surprise (PS)*, regarding the unforeseeable event. Each of these two conditions employs either a positive or a negative unforeseeable event (*High prize, Low prize*), resulting in a 2×2 between-subjects design. Figure 1 provides an overview of the timing in the experiment. We will provide a rationale for our choice of the four conditions after we spell out their details. Our reasoning behind the titles *Information surprise* and *Payment surprise* will also become apparent.

Under *IS* condition, the participants are presented with an urn, called the *original* urn, and informed that the urn contains balls with labels representing prizes measured in tokens. Each earned token is exchanged for €0.05 at the end of the experiment. The participants are told that: "*the urn contains two and only two prizes*". However, they are not told what these two prizes or their relative proportions are. Furthermore, we do not alert the participants that the composition of the urn might be changed by adding or removing balls from the urn.

Following the description of the urn, the participants observe a sequence of 20 physical draws from the original urn ($D^o_1 - D^o_{20}$ in Figure 1). The original urn contains 24 balls labeled '80' and 36 balls labeled '190'. As we already pointed out, this information is hidden from the respondents. All of the draws are made by a participant called experimental assistant, who is randomly selected for this task from the set of all participants. The outcome of each draw is revealed to all participants by the experimental assistant. Thus, everybody in a session observes the same sample. None of these 20 draws is payment-relevant. The only purpose of these draws is for the participants to gain information about the composition of the original urn.

After observing these 20 draws from the original urn, the participants are asked to provide estimates of the probabilities of the prizes they have been observing during these draws (subjective probabilities p^o_{80} and p^o_{190}) and to state their willingness-to-accept (smallest selling price) for the prospect to draw a prize from this urn (WTA^o). The procedures of eliciting probabilities and WTAs are described later in this section. The reported estimates of the two probabilities do not have to add up to 100%. The design does not force the sum of the estimates to be lower or greater than 100% either. We do not, however, explicitly ask for the respondents' estimates of observing a prize that they have not observed during the 20 draws. Thus, our design allows for an implicit estimation of a residual probability of outcomes that have not yet been observed. This contrasts with Experiment 2 where we explicitly ask for that probability.

Following the 20 draws and the reports of the two probabilities and WTA, a draw from the original urn is made by the experimental assistant. The outcome of this draw determines the potential payments for the reports of WTA^o. However, the draw is concealed from the participants when it is made. It is revealed only at the very end of the experiment when the final payment to the participants is displayed, provided that decision is selected for payment.

Subsequently, we bring out a new urn to the front of the experimental lab. One ball is then drawn from the new urn (D^n_1), revealing a new prize s to all participants. At this point in the experiment, the participants are informed that: "*This urn contains only the prize you are (about to be) shown.*" The value s of the new prize varies with the prize condition. In the *low* prize condition, the new urn contains 15 balls labeled '15', while in the *high* prize condition, it contains 15 balls labeled '375'. Although the participants know the value of all prizes in the new urn, they are not informed about how many balls are contained in the new urn. After revealing the value of the prizes in the new urn, we empty its contents into the original urn. We call this combined urn the *updated* urn. The participants are then asked to estimate the probabilities of each of the three prizes ($p^u_{80}, p^u_{190}, p^u_s$) and to state their WTA for the prospect to draw a prize from the updated urn (WTA^u). Following these reports, the experimental assistant draws a ball from the updated urn. This draw is concealed and is only revealed at the very end of the experiment, provided this decision is selected for payment.

Consider now the *PS* condition. Similarly to the *IS* condition, the participants are presented with the *original* urn, and are informed that the urn contains balls with labels that represent prizes. In contrast to *IS*, the participants are not told about the number of different prizes in the *original* urn. Similarly to *IS*, they are not provided with any information about the proportions of balls with any specific prize. In *PS* condition the respondents are informed that: "*at any point in the study new balls representing different tokens to what you have been observing so far may be added to this urn*". Thus, one might expect that some participants may incorporate this piece of information, which is not provided under *IS*, into their process of arriving at and reporting probabilities and WTAs. The participants subsequently observe 20 physical draws from the original urn ($D_1^o - D_{20}^o$ in Figure 1).

After observing the 20 draws from the *original* urn, the participants are asked to report their estimates of the probabilities of the prizes that they have been observing (subjective probabilities p_{80}^o and p_{190}^o) and to state their willingness-to-accept for the prospect to draw a prize from the urn (WTA^o). We denote with 'o' to emphasize that these values are elicited before any changes to the *original* urn are made. As for *IS*, the reported estimates of the probabilities are not restricted to add up to 100% or to be smaller or larger than 100%, allowing for calculation of an implicit residual probability.

Following the elicitation of these probabilities and WTA^o, we bring out a new urn to the front of the experimental lab. The participants are informed that: "*This urn contains new prizes. One such prize is the one you see. The urn contains no prizes similar to what you have been observing as a result of random draws from the other urn.*" The experimental assistant subsequently draws one ball from the new urn (D_1^n), revealing one new prize s to the participants. We do not reveal any other information about the contents of the new urn. As before, the value of the new prize s varies with the prize condition of the session. In the *low* prize condition, the new urn contains 15 balls labeled '15', while in the *high* prize condition it contains 15 balls labeled '375'. We then proceed to empty the contents of the new urn into the original urn, leading to the *updated* urn. The experimental assistant subsequently makes a draw from the updated urn. The outcome of the draw from the updated urn is used to potentially determine the payment for the report of WTA^o. However, the draw is concealed from the participants immediately after it is made. As for *IS*, it is revealed only at the very end of the experiment when the final payment to the participants is displayed, provided that the WTA^o report is selected for payment.

The participants are then again asked to estimate the probability of each prize ($p_{80}^u, p_{190}^u, p_s^u$) and to state their willingness-to-accept for the prospect to draw a prize from the urn (WTA^u). We denote with 'u' to emphasize that these values are elicited after an update to the urn occurs. Following these reports, the experimental assistant draws a ball from the updated urn. This draw is used to potentially determine the payment to the respondents, provided the WTA^u report is selected for payment at the very end of the experiment. Similarly to *IS*, this draw is concealed from the respondents and only revealed at the very end of the experiment if this decision is selected for payment.

The objective of our study is to induce an unforeseeable event and to examine individuals' reactions to it. To achieve this an event should (i) be unannounced and/or ruled out and (ii) have immediate payment consequences. Ideally, our treatment would have both of these characteristics. However, such a design could be interpreted as deceiving - we could be accused of explicitly or implicitly signaling that there would be no new event but then implementing the latter and making it payment-relevant. Due to this constraint, we designed our experiment to have two treatments, each of which has one and only one of these characteristics. In other words, we could not kill two birds with one stone and chose to kill two birds with two stones. In *IS* condition, the new event is unannounced while in *PS* condition the new event is payment-relevant because the payment is determined by the updated urn. Specifically, in contrast to *IS* condition, the respondents in *PS* condition are forewarned that new prizes may be added to the urn (so making the new event a bit less unforeseeable). Moreover, while in *IS* condition the first potentially payment-relevant draw is made from the original urn, the corresponding draw in *PS* condition is made from the updated urn. Our approach also allows us to test the robustness of results regarding reverse Bayesianism across different settings.

In addition to these two differences, the conditions *IS* and *PS* differ along two other dimensions. These two differences were implemented to make the new event in *IS* condition as unexpected as possible and to avoid misleading the respondents in *PS* condition as much as possible. The third and fourth differences pertain to the information about the compositions of the original and new urns, respectively, under the two conditions. Unlike in *IS* condition, we did not tell the respondents in *PS* condition that the original urn contains two and only two prizes. Otherwise, one could argue that we are sending a message that contradicts the possibility that the content of the urn may be changed or that we are trying to mislead the respondents. Even after making 20 draws from the original urn, the respondents in *PS* condition may expect to encounter a prize value that they have not yet observed. Thus, the possibility of drawing a new prize is conceivable in *PS* condition. Finally, the fourth difference is consonant with the third. The respondents in *IS* condition are told that the new urn contains only balls with the newly revealed prize. In contrast, under *PS* condition the possibility that the new urn may contain prizes other than the newly revealed prize is not ruled out.

The payment for the urn tasks is determined as follows. One item is randomly selected from the set of all reported probability estimates and the two WTAs. This item is played out and the resultant payoff is added to a participant's payment. We incentivize the reported probability estimates according to the Karni (2009) method.[4] The reports of *WTA* are incentivized using the BDM procedure (Becker et al., 1964). Both mechanisms induce truth telling and are robust to varying risk attitudes.

If one of the WTAs is selected to be payment-relevant, the computer makes a draw of a random price for the prospect between the smallest and largest prizes in the urn.[5] If the realization of the random price is such that the reported WTA exceeds the randomly generated price then the participant keeps the prospect and the payoff is determined by the hidden draw made during the experiment from the original urn in *IS* condition for WTA^o (see D_{21}^o in Figure 1); from the updated urn for WTA^u (see D_1^u in Figure 1); and from the updated urn for both WTA^o and WTA^u in *PS* condition (see D_1^u and D_2^u in Figure 1).

The experimental assistants who are selected to conduct the sample draws from the urns are not

[4]See in the first page of the experimental instructions following the heading "Likelihoods of events – Reporting and Earnings" for more details on how this was explained to the participants as well as further details on the method itself.

[5]Similarly to Isoni et al. (2011), these bounds are not communicated to the participants.

part of the experiment otherwise. They do not complete any of the tasks that the other participants perform and receive a fixed payment of €14, which is very close to the average earnings.

Once the urn task is completed, we elicit risk preferences using an incentivized Eckel and Grossman (2008) task. In this task, individuals pick one lottery from a set of binary lotteries. The lotteries in the choice set vary in terms of their expected values and variances with the chosen lottery revealing a participant's risk attitude. The lotteries chosen by the participants are "played out" at the end of the experiment and the earnings for these choices are added to the rest of the earnings of each participant.

Following the elicitation of risk preferences, the participants complete a short Raven Advanced Progressive Matrices (APM) test (Raven et al., 1998b,a). Raven's Progressive Matrices provide an effective non-verbal avenue to measure reasoning and general cognitive ability. In order to shorten the duration of this test, we follow Bors and Stokes (1998) in using 12 from the total of 36 matrices from Set II of the APM. Matrices from Set II of the APM are appropriate for adults and adolescents of higher average intelligence. Participants are allowed a maximum of 10 minutes. The participants are informed that two of these 12 matrices are selected at random for payment and that they will receive €1 for each correct choice. The sessions are concluded with some general demographic questions and a final screen informing the participants how their total earnings were calculated.

We include the experimental instructions in the appendix. The design was pre-registered at the AEA RCT Registry `https://www.socialscienceregistry.org/trials/3815`.

Implementation

Experiment 1 was conducted at the Alfred-Weber-Institute Experimental Lab at the University of Heidelberg and the Karlsruhe Decision & Design Lab (KD^2Lab) at the Karlsruhe Institute of Technology. The recruitment of participants took place via SONA systems for Heidelberg and ORSEE (Greiner, 2004) for Karlsruhe. A total of 344 participants participated in the experimental sessions.[6] The participants earned an average of €18.4, including a show-up fee of €4. The software used for the entire experiment was z-Tree (Fischbacher, 2007). The ethical approval for this design was granted by the Humanities and Social Sciences Research Ethics Sub-Co at the University of Warwick under DRAW Umbrella Approval (Ref: HSS 49/18-19, DR@Wsubmission ID: 485613261).

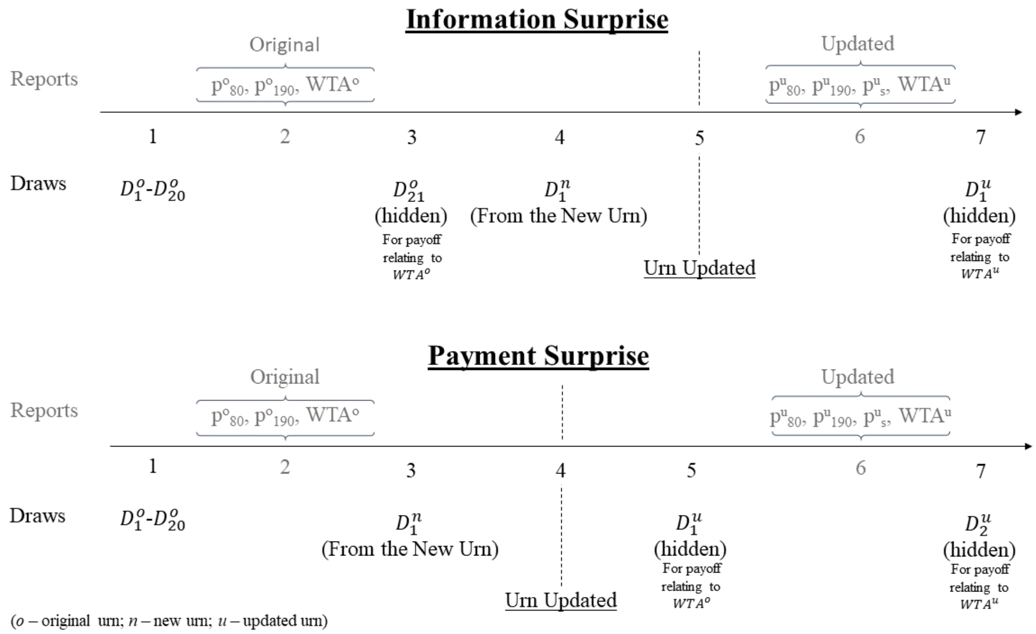
3.2 Results

3.2.1 Reverse Bayesianism

We start by testing whether belief updating is consistent with reverse Bayesianism. In our framework, reverse Bayesianism requires that the elicited probability ratios of the prizes in the original urn, namely the prizes of 80 and 190 tokens, remain unchanged after the original urn is updated. Formally, Hypothesis 1 for this experiement implies:

[6] We had 234 participants in the sessions at Heidelberg and 110 participants in the sessions at Karlsruhe. In Heidelberg, 46 participants were in the (*IS, low prize*) treatment, 58 in (*IS, high prize*), 59 in (*PS, low prize*), and 71 in (*IS, high prize*). In Karlsruhe, the numbers were 30, 17, 34, and 29, respectively.

Figure 1: The timing of the two main treatments



$$\Delta R = \frac{\hat{p}^o_{80}}{\hat{p}^o_{190}} - \frac{\hat{p}^u_{80}}{\hat{p}^u_{190}} = 0 \tag{1}$$

The information provided to the participants in all four treatments unambiguously reveals that the number of balls worth 80 tokens and the number of balls worth 190 tokens remain unchanged after the new urn is updated. For (*IS, low prize*) and (*IS, high prize*) treatments, we informed the participants that the new urn contains only the newly revealed prize. For (*PS, low prize*) and (*PS, high prize*) treatments, we told them that the new urn contains no prizes similar to what they have been observing as a result of random draws from the original urn. Thus, if a participant's beliefs are given by a singleton probability distribution, i.e. they are probabilistically sophisticated (Machina and Schmeidler, 1992), throughout the experiment, then $\Delta R = 0$ as long as that participant forms their beliefs on all of the information provided in the experimental instructions and as long as they do not expect that a disproportionate number of balls with prizes worth 80 and 190 tokens may be added to the urn they are facing.

Table 1 contains the results of Wilcoxon signed-rank test for all four treatments. We fail to reject the hypothesis in three out of the four treatments, both before and after correcting for multiple testing. For these three treatments, we indeed find a statistically significant null effect. Only in one treatment, (*PS, high prize*), we cannot reject the null hypothesis of reverse Bayesianism. Moreover, looking at the confidence intervals we note that the estimated normal distribution is such that 95% of participants in magnitude deviate very little from 0, even in the treatments where we reject the null hypothesis. Overall, we thus find strong evidence in support of reverse Bayesianism.

The ratio may remain unchanged under two scenarios. First, participants may simply not

Table 1: Average ratio changes before and after the urn is updated.

	Obs	z-statistic	p-value	p-value (corr.)	95%CI
IS, low prize	75	-0.89	0.375	1.000	$[-0.06, 0.05]$
IS, high prize	75	0.02	0.981	1.000	$[-0.06, 0.14]$
PS, low prize	93	-0.10	0.918	1.000	$[-0.06, 0.03]$
PS, high prize	100	2.55	0.011	0.043	$[-0.04, 0.05]$

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

change their estimates for all previously observed prizes. Table 2, column 7, shows that this is rarely the case in any of the treatments. Second, participants may change their estimates in a way that keeps the ratios constant. Notice that this is far from trivial for many constellations of estimates. Although we observe such instances (column 4), overall this is less prevalent than the instances when participants either increased or decreased the ratio, however even when the ratio changes, as we already noted from table 1, most participants that change their ratios do so only slightly. That is, the overall null effect in support of reverse Bayesianism in the first three conditions derives from, if any changes, very small ones.

Figure 2 shows the distribution of absolute changes in the ratios for the participants of different degree of cognitive ability. The data is pooled over all four treatments. Our analysis reveals that the absolute differences of the ratios tend to be concentrated around zero, independently of the participants' cognitive ability. That is, the change in the ratio is relatively small even for those participants who updated their beliefs to a different ratio. Moreover, we do not find a statistically significant relation between cognitive ability and the absolute ratio differences (see Appendix A, Table 12, OLS). That is, behavior is very similar for higher and lower cognitive ability participants, and thus unlikely to be due to errors caused by any lack of understanding.

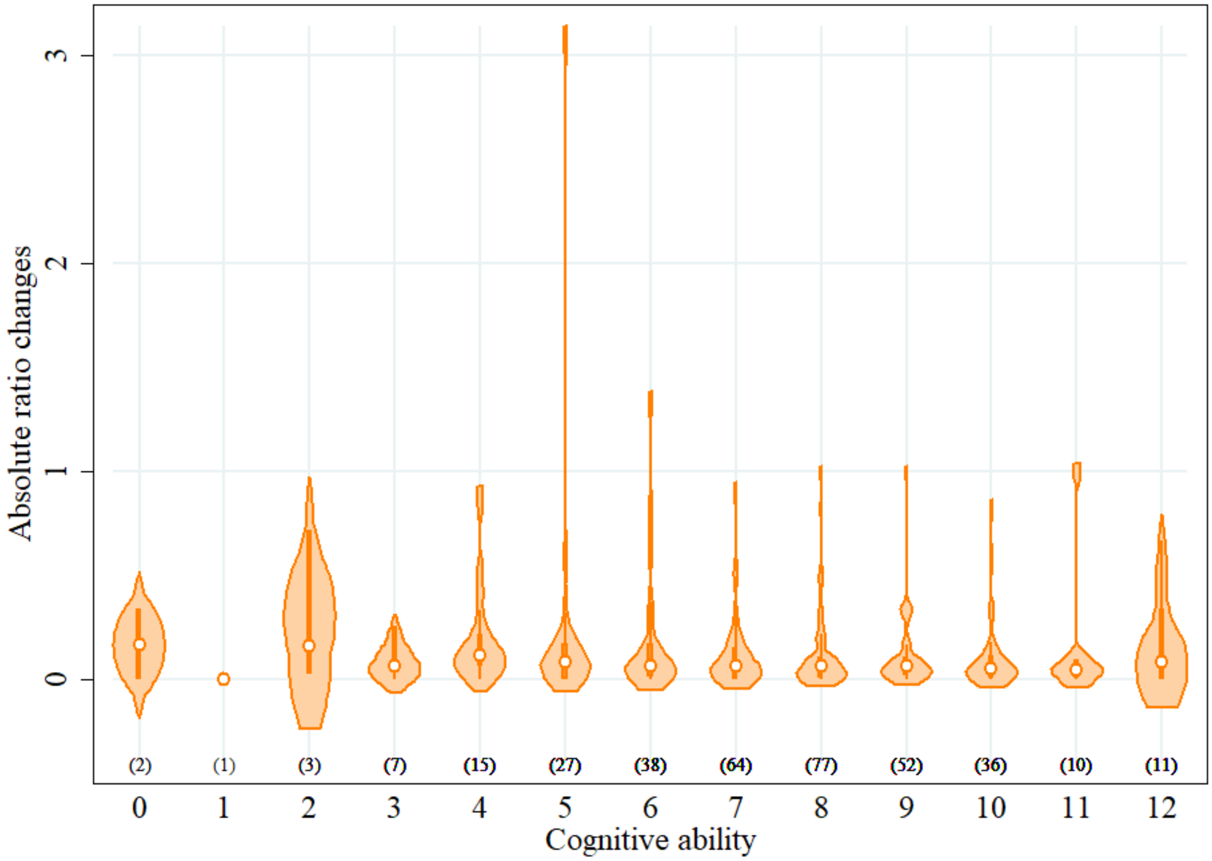
Thus, overall we find clear evidence in support of reverse Bayesian updating. Participants seem to aim to hold their ratios approximately constant after an unexpected event occurs.

Table 2: Increase or decrease of ratios before vs. after observing updating the urn.

	Increased	Decreased	Const ratio	p-value	p-value (corr.)	Unchanged Est.
IS, low prize	23	29	23	0.488	1.000	1
IS, high prize	32	31	12	1.000	1.000	1
PS, low prize	37	33	23	0.720	1.000	0
PS, high prize	61	29	10	0.001	0.004	4

Matched pairs sign test i.e. testing that the median of differences is zero, p-values corrected by Bonferroni-Holm procedure. 'Const ratio' denotes participants holding their ratio constant, 'Unchanged Est.' denotes that participants did not change their estimates.

Figure 2: Violin plot of the change in ratios before and after observing a new color depending the ratios remained constant or not.



3.2.2 Residuals and Valuations

The experiment yields a number of additional interesting findings. First, we focus on the estimates of residual probabilities provided by the participants and test Hypotheses 2 and 3. Considering Hypothesis 2, the participants in the *IS* condition were informed that the original urn contains only two possible prizes, and that in the updated urn only one extra prize was added. If the respondents took this information into account the set of residual consequences x, as defined in section 2, is empty, thus $\hat{p}_x^o \neq 0$ and $\hat{p}_x^u \neq 0$ would likely materialize only due to individual idiosyncratic errors. In contrast, since in the *PS* condition we informed the participants about the possibility of adding new prizes and we did not state that the updated urn contains only one extra prize, $\hat{p}_x^o > 0$ and $\hat{p}_x^u > 0$ could occur as a result of an expectation of further yet unknown events. Thus, one could expect that the respondents in the *PS* condition are more likely to assign a strictly positive probability to encountering a prize that they have not seen before.

Table 3 shows that the hypothesis $\hat{p}_x^j = 0$ cannot be rejected for the *IS* condition, both for the original and in the updated urn, thus giving support to Hypothesis 2. However, even in the *PS* condition, the hypothesis $\hat{p}_x^j = 0$ cannot be rejected for the original urn and the updated urn in cases of (a perhaps less salient) high prize surprise. The hypothesis that $\hat{p}_x^j = 0$ can be rejected in the updated urn in the *PS* low prize treatment. Moreover, from columns 4 and 5 of table 3, we can observe that in the updated urn of the *PS* condition there is a fair number of participants for which $\hat{p}_x^u > 0$.

It is arguable that in the original urn participants will face an unforeseeable event, while for the updated urn, participants could in principle expect a new urn with new prizes appearing. We can then conclude the analysis following the test of hypothesis 2 by saying that participants do not seem to expect unforeseeable events, but there is some evidence suggesting that some of them may expect a new event when such events become foreseeable.

Table 3: Residuals different from 0.

	p-value	q-value	$\hat{p}_x = 0$	$\hat{p}_x > 0$	$\hat{p}_x < 0$
Original, IS low prize	0.993	1.000	74	1	1
Original, IS high prize	0.314	1.000	71	3	1
Original, PS low prize	0.317	1.000	92	0	1
Original, PS high prize	0.549	1.000	90	6	4
Updated, IS low prize	0.251	1.000	61	10	5
Updated, IS high prize	0.228	1.000	65	7	3
Updated, PS low prize	0.004	0.028	74	16	3
Updated, PS high prize	0.146	1.000	84	11	5

We turn now to Hypothesis 3. It seems plausible that awareness of encountering unexpected events increases after a new urn is revealed and its contents are added to the original urn. In the *IS* condition, we announce that the new urn contains only one prize, which rules out the possibility of existence of any other prizes than the prize that was revealed to the participants. In the *PS* condition, the possibility of another prize is not explicitly excluded. Indeed, from table 4, we note that the hypothesis that $\hat{p}_x^o = \hat{p}_x^u$ cannot be rejected in the *IS* treatments. However, there is again some evidence for $\hat{p}_x^o(PS) \neq \hat{p}_x^u(PS)$, when the surprising event is a low prize.

Table 4: Differences between residuals before and after the surprise, $\Delta\hat{p}_x = \hat{p}_x^u - \hat{p}_x^0 \neq 0$.

	p-value	q-value	$\Delta\hat{p}_x = 0$	$\Delta\hat{p}_x < 0$	$\Delta\hat{p}_x > 0$
IS low prize	0.285	1.000	8	40	28
IS high prize	0.560	1.000	15	30	30
PS low prize	0.019	0.075	12	31	50
PS high prize	0.951	1.000	10	40	50

Given our design, awareness of encountering an unforeseen event and the way an unforeseen event is experienced will affect the stated $WTAs$. It follows from Karni and Vierø (Karni and Vierø, 2013, 2017) that two key factors are at play in the participants' evaluation of uncertain prospects. The first characteristic pertains to the participants' updated beliefs; how much of the probability weight is shifted from the known prizes to the newly observed and yet unobserved prizes. The second pertains to the participants' attitude towards the unknown; whether and how much they like or dislike the unknown. To determine the relative importance of these channels, we compare the elicited willingness to accept measures before and after the urns are updated, for both *IS* and *PS* conditions and both levels of the new prize. Table 5 reveals that $WTA^o(PS) > WTA^o(IS)$ in both high and low prize conditions. As it is reasonable to expect, the $WTAs$ for the updated urn are lower for the low prize than for the high prize, and again the *PS* condition elicits higher valuations; the latter effect is, however, not significant.

Regression analyses with controls for gender, cognitive ability and the observed sample, confirm the effect of the *PS* treatment and of the high prize in the updated urn *WTA* (see Tables 13 and 14 in Appendix A). Overall, it seems that the more uncertain situation in condition *PS* elicits higher valuations. That is, in the context of unforeseen events, *hope* seems to dominate *fear* (Viscusi and Chesson, 1999). As a caveat, we note that WTA measurement in the context of uncertainty and ambiguity has been found to elicit relatively higher valuations for more uncertain prospects (Trautmann et al., 2011; Trautmann and Schmidt, 2012). The selling-price context seems to induce decision makers to focus on the potentially forgone benefits from selling a highly uncertain prospect. This effect re-emerges here.

Table 5: WTA for a draw from the urn by treatment.

Original urn: $\boldsymbol{WTA^o}$

	IS	PS	Diff	p-value
Low prize	110.39	138.47	-28.08	0.008
High prize	110.48	134.81	-24.33	0.002

Wilcoxon signed-rank test.

Updated urn: $\boldsymbol{WTA^u}$

	IS	PS	Diff	p-value
Low prize	86.45	96.70	-10.25	0.074
High prize	153.53	178.25	-24.72	0.160

Wilcoxon signed-rank test.

4 Experiment 2

4.1 Design

Experiment 2 tests how individuals perceive uncertainty and update their beliefs in light of new events which are foreseeable, but potentially unforeseen. Each participant in Experiment 2 individually draws a sample of 30 colored marbles from a virtual urn in each of four different tasks. Our treatments correspond to the order in which the participants faced these tasks. The urns in each of the four tasks contain a total of 100 virtual marbles, which was known to the participants. No other information about the composition of the urns is revealed to the participants prior to them making the draws from the urns. Additionally, the exact colors for each task are randomized at the participant level, i.e., each participant observed a different combination of colors in a certain task.

Each participant is randomly allocated to one of two treatments. In the *two colors* treatment, the urn in the first task contains only two colors. In the *four colors* treatment, the urn in the first task contains four colors in total. The purpose of this design is to test if encountering a larger number of different colors in the first task increases their awareness that further surprises might be possible in subsequent tasks. The compositions of the urns in the second, third and fourth tasks are the same across the two treatments. The urn in the second task contains three colors, the urn in the third task contains two colors and the urn in the fourth task contains four colors. Table 6 provides information on the exact compositions of the urns in the four tasks.

Table 6: Number of different colors in the tasks.

	Task 1		**Task 2**	**Task 3**	**Task 4**
	Two colors	Four colors			
color 1	55	40	53	75	48
color 2	45	28	35	25	28
color 3		20	12		12
color 4		12			12

The urns with three and four colors contained a comparatively small number of some of the colors. This ensured that the likelihood of encountering a new and surprising outcome even after sampling several times was relatively large for these urns.

After each sample draw, the drawn marble is presented on a participant's screen, both with a graphical depiction of a marble of that specific color and the name of the color. In addition, the outcomes of all previous draws are depicted in a small caption at the bottoms of the participants' screens. This provides the participants with a full overview of the past draws and mitigates the effects of memory limitations.

After each draw, the participants are asked to report their estimates of the contents of the urn. Specifically, they are asked to state their estimate of (i) the number of marbles of the color they just drew; (ii) the number of marbles of each other color they have previously drawn in this task; and (iii) the number of marbles of yet unobserved colors.

The third item provides us with a residual probability assigned by the respondents to any conceivable color not yet observed during the draws. Thus, in contrast to Experiment 1, the residual probability is elicited explicitly in Experiment 2. The participants submit their estimates by entering integer numbers between 0 and 100 into form fields on their screens. We provide them with one individual form field for each color drawn up to that point, as well as with one form field for their estimate of the number of marbles of yet unobserved colors. This design allows us to trace how the respondents' estimates and ratios of these estimates are adjusted once unforeseen information becomes available. In addition, we can track how the estimates of likelihoods of yet unobserved outcomes evolve over the sampling process. To make the submission of estimates easier, we additionally provide respondents with buttons that allow them to fill in their last estimates for a color and "plus" and "minus" buttons to increase/decrease these estimates by one per click.

We use the Karni (2009) method to ensure that the participants are incentivized to submit their estimates truthfully. At the end of the experiment, for each of the four tasks, one of the 30 sample draws is randomly selected and one item from the set of reported estimates for that draw is randomly selected (this could involve an estimate of yet unobserved colors). The payment mechanism is implemented for that reported estimate. Participants can earn £6 or nothing from each task depending on the outcome according to the incentivisation method.[7]

In addition to the hypotheses presented in Section 2, we additionally test the following:

Hypothesis 4.

(a) The participants expect fewer unforeseen events the more samples they draw:

$$\#draws \to 30 \Rightarrow \hat{p}_x \downarrow$$

(b) The probability assigned to yet unobserved outcomes in tasks 2, 3 and 4 decreases faster for the participants in the two colors treatment than for the participants in the four colors treatment

Following some recent insights on the link between cognitive ability and rational behaviour (e.g. Gill and Prowse, 2016; Alaoui and Penta, 2016) we also test two hypotheses related to cognitive ability.

Hypothesis 5. *The participants with higher cognitive ability exhibit fewer deviations from reverse-Bayesianism.*

Hypothesis 6. *The participants with higher cognitive ability expect yet unobserved outcomes up to a later point in the sampling process than the participants with lower cognitive ability.*

We include the experimental instructions in the appendix. The design was pre-registered at the AEA RCT Registry `https://www.socialscienceregistry.org/trials/5499`.

[7]See in the third page of the experimental instructions for this experiment under the heading "Getting paid for good predictions" for more details on how this was explained to the participants as well as further details on the method itself.

Implementation

Experiment 2 took place at the Behavioral Science Lab at the University of Warwick. The recruitment was conducted with the DRAW (Decision Research at Warwick) system, based on the SONA recruitment software. A total of 174 participants participated in the experiment, 89 in the *two colors* treatment and 85 in the *four colors* treatment. The average payment was £16.87, including a show-up fee of £3. The software for the experiment was programmed in otree (Chen et al., 2016). Ethical Approval for this design was granted by the Humanities and Social Sciences Research Ethics Sub-Co at the University of Warwick under DRAW Umbrella Approval (Ref: HSSREC 104/19-20, DR@W submission ID: 514470520).

4.2 Results

4.2.1 Reverse Bayesianism

We first test whether behavior is consistent with reverse Bayesianism, which corresponds to Hypothesis 1. We examine the difference in the ratios of previously observed colors directly before and directly after observing a new color. The ratios are defined for pairs of colors that have already been observed and on the basis of the relative magnitudes of the estimated likelihoods of these two colors immediately before the new event is observed. Specifically, we define $\hat{p}^o_H$ as the estimate of the likelihood for the color that is considered by a participant to be more likely and $\hat{p}^o_L$ as the estimate for the color that is considered to be less likely. Both of these estimates are for the sample draw right before the third color is observed for the first time. We also define $\hat{p}^u_H$ and $\hat{p}^u_L$ as the estimates of the likelihoods of these two colors in the sample draw when the third color is first observed. Specifically, we test:

$$\Delta R^3 = \frac{\hat{p}^u_H}{\hat{p}^u_L} - \frac{\hat{p}^o_H}{\hat{p}^o_L} = 0,$$

For the belief update after observing a fourth color (having seen three colors before) we have three ratios to consider. We define $\hat{p}^o_H$, $\hat{p}^o_M$ and $\hat{p}^o_L$ as the estimates for the color considered most likely, second most likely, and least likely, of the three colors that have already been observed, in the sample draw right before the fourth color is observed for the first time. We define $\hat{p}^u_H$, $\hat{p}^u_M$ $\hat{p}^u_L$ as the estimates for these three colors in the sample draw when the fourth color is first observed. We test three relationships:

$$\Delta R^4_1 = \frac{\hat{p}^u_H}{\hat{p}^u_M} - \frac{\hat{p}^o_H}{\hat{p}^o_M} = 0, \quad \Delta R^4_2 = \frac{\hat{p}^u_M}{\hat{p}^u_L} - \frac{\hat{p}^o_M}{\hat{p}^o_L} = 0, \quad \Delta R^4_3 = \frac{\hat{p}^u_H}{\hat{p}^u_L} - \frac{\hat{p}^o_H}{\hat{p}^o_L} = 0.$$

Table 7 shows that for all nine tests that we conduct, there is no significant change in the ratios after controlling for multiple tests.[8] Even when not controlling for multiple testing, the ratio does not exhibit a statistically significant change in seven of the nine instances. Finally, from the 95% confidence intervals in column 5, we notice that ratios do not change substantially

[8] The urn in task 3 contains only two colors. Hence, the third outcome surprise is not possible and, consequently, we do not analyze it in this case.

among participants in the estimated normal distribution, especially when they only have to keep one ratio constant, like in ΔR^3 (when only a 3^{rd} color is observed). As noted in the analysis for Experiment 1, the finding that ratios do not significantly change in aggregate can derive from (i) participants holding their estimates unchanged, (ii) holding ratios constant, or that (iii) some participants increase and some decrease their ratios, canceling out on average. Table 8 indicates that, in contrast to Experiment 1, a large number of participants hold their estimates unchanged. A possible explanation for participants in Experiment 2 not changing their estimates is the provision of a button to fill-in their previous estimate to simplify the dynamic task for the participants. However, there is also a substantial number of participants who hold their ratio constant, while adjusting the separate probability estimates. Again, holding the ratio constant while adjusting the separate estimates, especially after the fourth color is observed, is far from trivial. It indicates that participants actively aim to keep ratios constant even when changing their separate estimates. At the same time, a substantial share of the participants change their ratios, but there is no systematic effect in the way they do it: after correcting for multiple testing, there are no significant differences in the deviations from constant ratios in the increasing or decreasing directions; seven of the nine uncorrected tests support constant average ratios.

Figure 3 illustrates that, as in Experiment 1, cognitive ability does not have a significant effect on ratio deviations (see also Appendix B, table 15). Thus, there is no empirical evidence supporting Hypothesis 5. Importantly, the absolute changes in the ratios are also unaffected by the participants' expectations of further surprises, that is, whether they hold non-zero residual probabilities or not (Appendix B, table 16).

Table 7: Average ratio changes before vs. after observing a new color.

	Obs	z-statistic	p-value	p-value (corr.)	$95\%CI$
Task 1 ΔR^3	85	-1.37	0.172	1.000	$[-0.10, 0.29]$
Task 1 ΔR_1^4	84	-0.55	0.584	1.000	$[-0.79, 0.22]$
Task 1 ΔR_2^4	84	-2.13	0.033	0.296	$[-1.27, 0.04]$
Task 1 ΔR_3^4	84	-1.00	0.315	1.000	$[-0.52, 0.33]$
Task 2 ΔR^3	169	-2.63	0.008	0.076	$[-0.31, 0.01]$
Task 4 ΔR^3	173	-0.65	0.517	1.000	$[-0.26, 0.06]$
Task 4 ΔR_1^4	164	-0.05	0.962	1.000	$[-0.07, 0.25]$
Task 4 ΔR_2^4	163	-0.07	0.946	1.000	$[-0.19, 0.69]$
Task 4 ΔR_3^4	163	-0.15	0.883	1.000	$[-0.14, 0.27]$

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

4.2.2 Analysis of the Residuals and the Dynamics of the Beliefs

We now turn to the estimates of the residual probabilities. Figure 4 depicts the evolution of the average residual $\hat{p}_x$ over the 30 sample draws for each task for both treatments. We find that for both treatments the average residual starts at a relatively high level and decreases quickly as more

Table 8: Increase or decrease of ratios before vs. after observing a new outcome.

	Increased	Decreased	Const ratio	p-value	p-value (corr.)	Unchanged Est.
Task 1 ΔR^3	16	29	40	0.072	0.652	26
Task 1 ΔR^4_1	19	21	44	0.875	1.000	31
Task 1 ΔR^4_2	16	31	37	0.040	0.360	32
Task 1 ΔR^4_3	16	23	45	0.337	1.000	35
Task 2 ΔR^3	35	59	75	0.017	0.155	46
Task 4 ΔR^3	45	50	78	0.682	1.000	44
Task 4 ΔR^4_1	50	57	57	0.562	1.000	36
Task 4 ΔR^4_2	54	60	49	0.640	1.000	33
Task 4 ΔR^4_3	43	47	73	0.752	1.000	37

Matched pairs sign test i.e. testing that the median of differences is zero, p-values corrected by Bonferroni-Holm procedure. 'Const ratio' denotes participants holding their ratio constant, 'Unchanged Est.' denotes that participants did not change their estimates.

draws are made. Note how even after the 30^{th} sample draw the average residual is well above 0. In figure 8 of the appendix, we present the distribution of these residuals; about one third of the participants expect further colors even after the 30^{th} sample draw. This gives support to Hypothesis 2.

For the sake of exposition it is better to postpone the analysis of Hypothesis 3 below, and consider now Hypothesis 4. We already observed how in figure 4 residuals on average monotonically decrease in sample draws. Simple correlations (across the whole set of participants) between the number of marbles already drawn from the virtual urn and the stated residuals are negative for all tasks and across both treatments ($\rho < -0.311$, Pearson correlation coefficient), and thus support the first part of Hypothesis 4. However, there are no significant differences between the two treatments in the overall shape of the curve (Kolmogorov-Smirnov test, all $p - values \geq 0.994$), thus rejecting the second part of Hypothesis 4 (differences in awareness of unexpected events). This indicates that participants do not adapt their estimation of residuals in later tasks in response to encountering more possible outcomes in the first task. This finding is intriguing. On the one hand, encountering more possible outcomes in the first task could raise the participants' awareness that the urns might contain more outcomes than initially expected, which was our prediction. On the other hand, as participants have no information ex ante and as task 1 does not provide direct information on subsequent tasks, the null effect we find might be perfectly rational.

Considering now Hypothesis 3 (residual changes after observing new events), there is a negative correlation between the number of colors already observed by the participants and their residuals ($r_s < -0.272$, Spearman correlation coefficient). That is, the more surprises the participants observe the fewer future surprises they anticipate. Looking at the residual directly before and after a new color is observed, a significant drop in the subjective residual probability is observed in 8 out of 9 instances – see table 9.

Figure 3: Violin plot of the change in ratios before and after observing a new color depending on the Raven score.

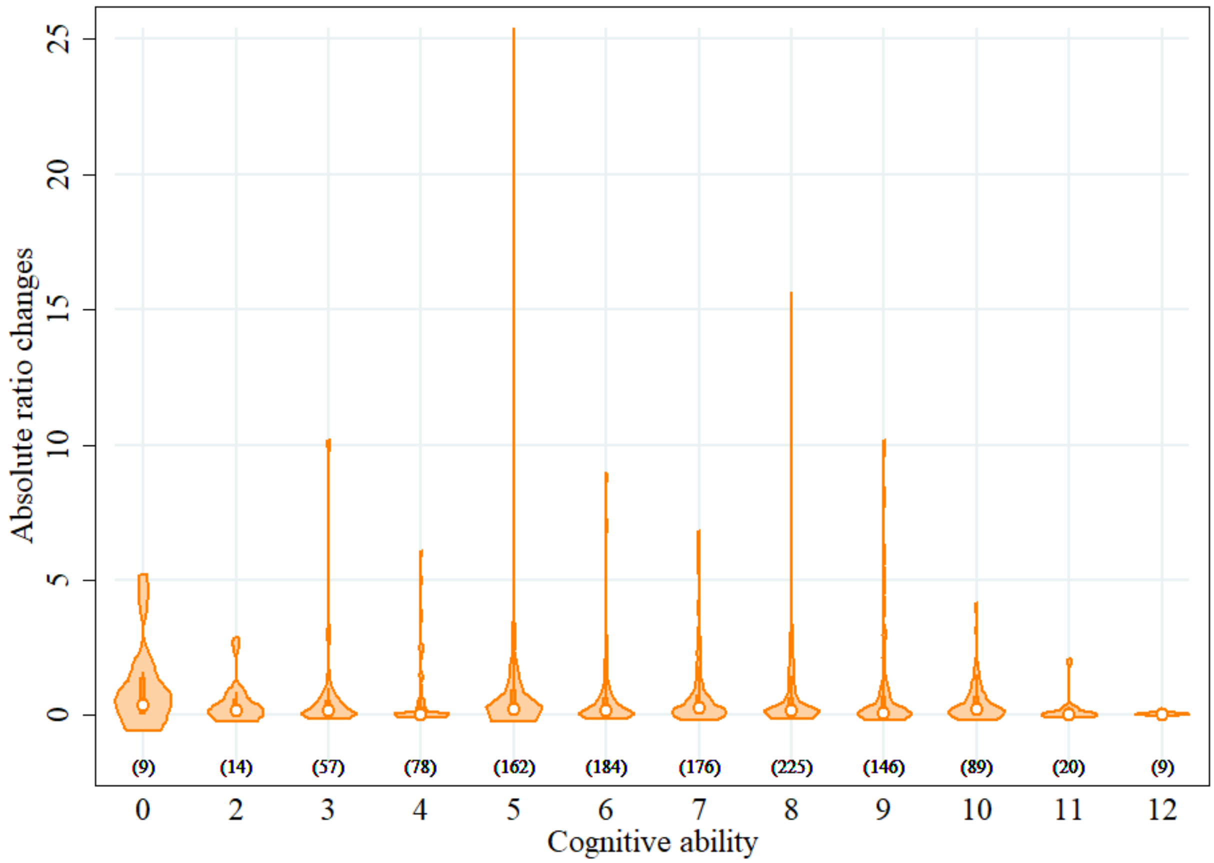


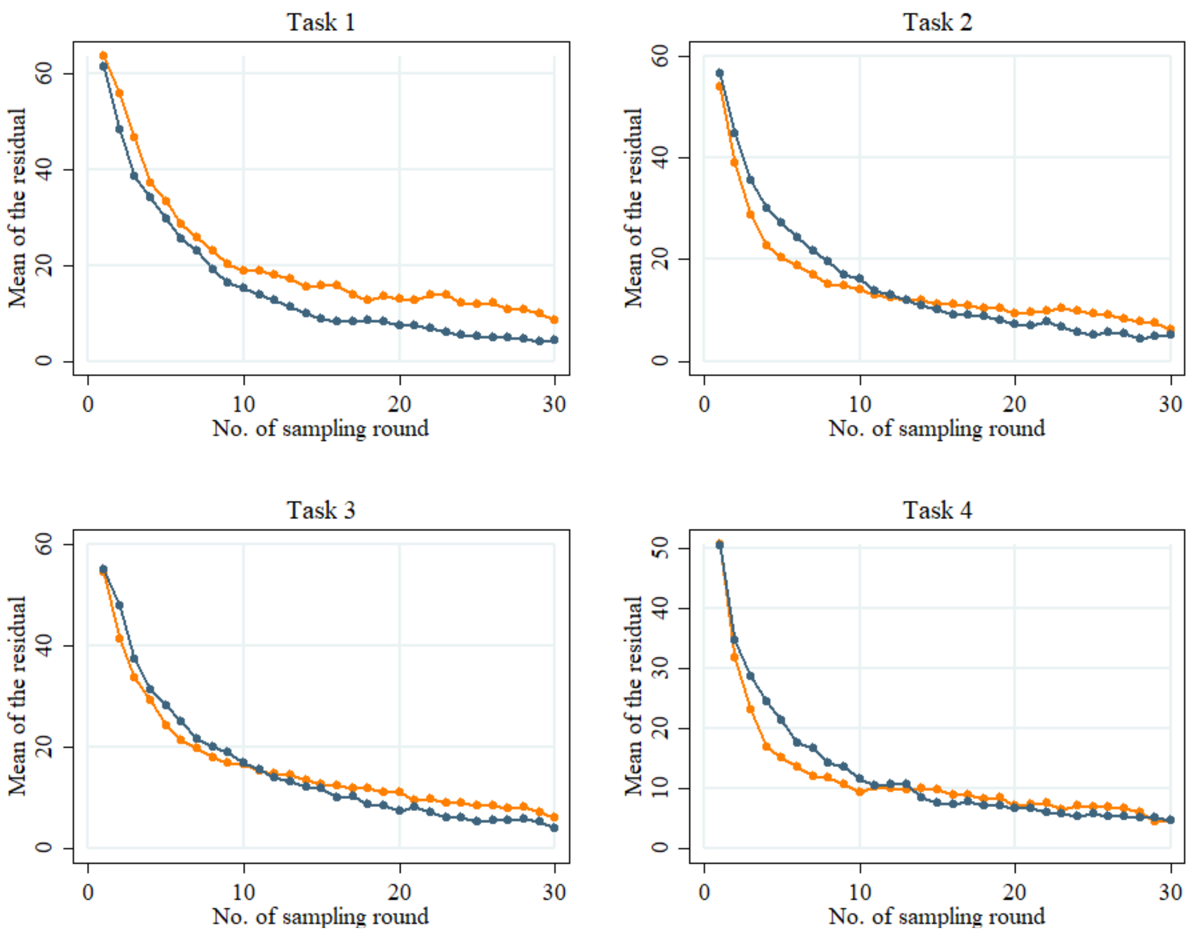
Table 9: Increase or decrease of residuals before vs. after observing a new color.

	Increased	Decreased	Constant	p-value	p-value (corrected)
Task 1 - $\hat{p}_x^2$	9	144	21	0.000	0.000
Task 1 - $\hat{p}_x^3$	16	56	13	0.000	0.000
Task 1 - $\hat{p}_x^4$	25	38	21	0.130	1.000
Task 2 - $\hat{p}_x^2$	3	149	22	0.000	0.000
Task 2 - $\hat{p}_x^3$	26	84	59	0.000	0.000
Task 3 - $\hat{p}_x^2$	3	149	22	0.000	0.000
Task 4 - $\hat{p}_x^2$	6	143	25	0.000	0.000
Task 4 - $\hat{p}_x^3$	27	94	53	0.000	0.000
Task 4 - $\hat{p}_x^4$	27	57	80	0.001	0.013

Matched pairs sign test, p-values corrected by Bonferroni-Holm procedure.
The i in $\hat{p}_x^i$ represents which number of new color has been observed.

To summarize so far: (1) drawing a larger sample decreases the residual as overall more precise

Figure 4: Development of the residuals over the sampling process.



Note: Orange line shows residuals of *two outcomes* treatment, blue line shows residuals of *four outcomes* treatment.

information on already observed colors is available; (2) observing more colors, ceteris paribus decreases the residual, perhaps because participants feel that the space of so far unobserved events shrinks the more colors are observed.

In order to better understand to which degree these two factors impact the residual, we estimate a random effects model of the residual on the number of draws and the number of observed colors, controlling for different demographic factors and the task numbers. Table 10 presents the results of this analysis. Both factors independently have a negative and significant impact on the size of the residual. The impact of the number of outcomes is roughly 11 stronger than the one of sample draws. Furthermore, always looking at table 10, it is interesting to note that a higher cognitive ability leads to a smaller residual in contradiction to Hypothesis 6. However, since it is not obvious what "optimal" residual an individual participant should have in this experiment, it is not possible to assess whether it is reasonable to observe this relationship. Finally, we run two robustness checks of the estimations presented in table 10. First, the coefficients and their significance remain similar when a fixed-effects instead of a random-effects model is used (Appendix B, table 17). Second, conducting the random effects panel analysis solely upon the latter half of the sampling process (only for observations after sampling round 15), leads to a smaller yet still significant negative coefficient for the number of draws (Appendix B, table 18). The coefficient for the number of colors observed is now not significant, possibly due to a relatively small number of new colors being observed in later rounds of the sampling process. Taken together, these results indicate that even in later stages of the sampling process, participants on average reduce their residual with every additional draw.

Table 10: Random Effects Estimator: Relation between sample draws and the residuals

Size of the residuals	
Num. draws	-0.746**
	(0.048)
Num. colours observed	-8.402**
	(0.537)
Cognitive ability	-2.624**
	(0.547)
Four colours first	-0.914
	(2.509)
Constant	77.907**
	(11.212)
Observations	19,800
Subjects	165

* $p < 0.05$; ** $p < 0.01$

Clustered standard errors in parentheses

4.3 Bayesian and Reverse Bayesian Rationality

As shown in subsection 4.2.1, the participants' behavior is consistent with reverse Bayesian reasoning. Does this mean that our participants are in general very rational Bayesian updaters? In order to test this, we study how observing a new color affects the sum of the new residual and the estimate of the new color. Technically, before actually observing the new color, its estimate should be included in the estimate of the event *any other color*. Observing a new color can be viewed as unpacking the estimate of the likelihood of yet unobserved colors into two new estimates, an estimate for the new color and another estimate for the yet unobserved colors. That is, the sum of the estimate of the new color and the new residual should equal the previous residual. Tversky and Koehler (1994); Sonnemann et al. (2013) find that the sum of such two unpacked estimates violate this principle; the estimate of the new color and the new estimate of any yet unobserved color, often exceeds the 'packed' estimate.

We replicate this violation of rationality in our data. We define $\hat{p}^u_{Sum} = \hat{p}^u_x + \hat{p}^u_C$ as the sum of the new residual and the estimate for the newly observed color C. As discussed, $\hat{p}^u_{Sum}$ should equal the previous-round residual $\hat{p}^o_x$. Table 11 tests if the ratios $\frac{\hat{p}^u_{Sum}}{\hat{p}^o_x}$ are different from 1. This is clearly the case for all instances of updates. Figure 5 illustrates this for when the second color is observed.[9] There is also a significant positive correlation between the number of colors observed and the unpacking ratio (Kendall's $\tau_A = 0.667, \tau_B = 0.785, p = 0.0095$). That is, the unpacking effects get more pronounced the more colors are already observed. This illustrates that our participants are indeed prone to violations of rational updating principles in the current context. This makes the strong evidence for reverse Bayesianism all the more remarkable.

Figure 5: Residual and the estimate of the new color before and after observing the new color, second color.

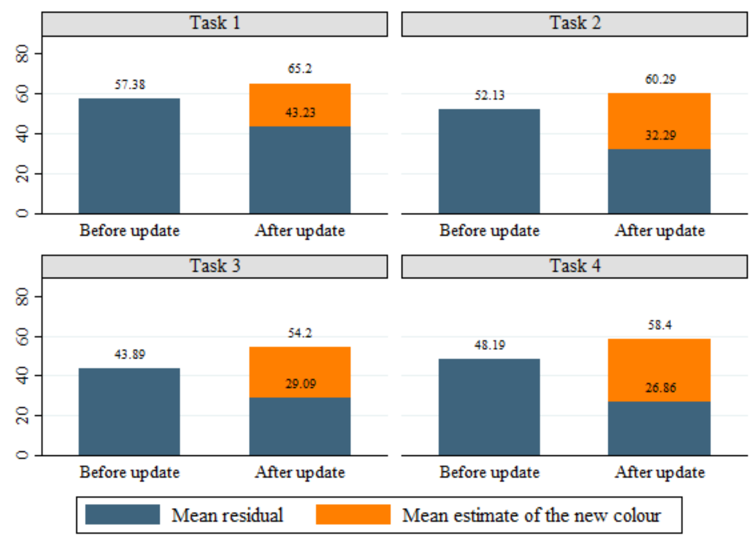


Table 11: Unpacking the residual after observing a new color.

	Unpacking factor	Obs	z-statistic	p-value	p-value (corr.)
Task 1 $\hat{p}^2\_Sum$	1.23	174	7.98	0.000	0.000
Task 2 $\hat{p}^2_{Sum}$	1.29	174	6.78	0.000	0.000
Task 3 $\hat{p}^2_{Sum}$	1.48	174	8.76	0.000	0.000
Task 4 $\hat{p}^2_{Sum}$	1.25	174	7.78	0.000	0.000
Task 1 $\hat{p}^3_{Sum}$	1.46	85	6.90	0.000	0.000
Task 2 $\hat{p}^3_{Sum}$	1.65	169	9.91	0.000	0.000
Task 4 $\hat{p}^3_{Sum}$	2.02	174	10.06	0.000	0.000
Task 1 $\hat{p}^4_{Sum}$	2.32	84	6.85	0.000	0.000
Task 4 $\hat{p}^4_{Sum}$	2.23	164	10.24	0.000	0.000

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

[9]See Figures 9 and 10 in Appendix B for an illustration of this for when the third and fourth colors are observed.

5 Discussion

Many studies assess how decision makers update beliefs about known events in empirical or experimental decision situations. We focus on new and more or less unforeseeable events. Different branches of literature offer various prescriptions on how to integrate information about unforeseen events into beliefs. In their seminal work, Karni and Vierø (2013; 2017) propose that it is optimal to integrate unforeseen events into a probability distribution such that the ratio of previously observed events stays constant. Our results provide evidence that also from a descriptive perspective, this reverse Bayesianism is compelling. This stands in sharp contrast to many studies on Bayesian updating, which often find behavior in experiments violating theoretical prescriptions (Charness and Levin, 2005; Charness et al., 2007; Holt, 2009). In other words, our results imply that reverse Bayesianism is compelling, both, *normatively* and *descriptively*. This holds true both, in situations involving reasonably unforeseeable events (Experiment 1), as well as unknown but foreseeable events (Experiment 2).

These implications are intriguing. The space of possible events is continuously expanding in many decision environments. Examples like the financial crisis of 2007 and the COVID-19 pandemic highlight the difficulty to react to novel events appropriately. It is thus important to pay special attention to the possibility of not knowing relevant events beforehand. While our experiments can not speak to the optimal reactions towards an unforeseen event, they show that decision makers have the capacity to reconcile new information optimally with their existing beliefs. Notably, this is irrespective of whether participants started with not explicitly expecting a surprise (Experiment 1) or if they were asked about such a belief and supplied a positive estimate (Experiment 2). Furthermore, it is noteworthy that participants in Experiment 2 reduced their residual the more draws they observed. This indicates an inclination towards complacency within the experiment, if less new outcomes are discovered. Testing this tendency in more applied scenarios and studying how it interacts with precautionary measures could be an interesting extension to our findings.

Our findings for Experiment 1 also indicate that participants exhibit a higher WTA when aware of the possibility of further, unknown surprises. In the interpretation of Viscusi and Chesson (1999), we observe that people seem to be more hopeful rather than fearful towards unknown future events. This is interesting with respect the ambiguity literature (Trautmann and van de Kuilen, 2015). One interpretation is that low probabilities are attached to unforeseen events (see also Experiment 2), and that this is an instance of ambiguity seeking. Indeed, the Trautmann and van de Kuilen (2015) review reports predominantly ambiguity seeking for low probability gain prospects like the ones used in Experiment 1.

In Experiment 2 (and implicitly in Experiment 1) we used the residual estimate as a catch-all way of encoding the participants beliefs about all remaining possible events. This does however not give us a clear description of what exactly participants expect in the future, instead it is a "[...]Black Box, a residual of unknown content." (Shackle, 1992, p. 23). For example, a participant stating a positive residual would expect a surprise in the future. Say she expects the urn to also contain blue and red marbles. Suddenly observing a purple marble could present her with an unforeseen event. Further studies could try to elicit an exhaustive list of expected events from the participants or even try to use a completely non-distributional approach to assess uncertainty (Shackle, 1992).

Furthermore, the surprises in our experiments might still be considered to be easily comprehensible. A new prize in an urn of prizes, a new color in a box of colored marbles, can be surprising and unexpected, but is simple to integrate in a existing belief structure after its first occurrence. A next step could be to test to which degree the principle of reverse Bayesianism extends to more complex events, where it is less clear which shape a surprise might take and how to reconcile it with existing beliefs. This could help us to better understand how belief systems are affected by completely rare and novel events.

Finally, our results cannot resolve the question whether participants act as-if they are reverse Bayesian or if this behavior is a deliberate rationalization. On the one hand, some participants did not alter their estimates at all after observing a new outcome. On the other hand, many did, and still provided belief ratios that were wither constant or deviated little from the previous ratio. Both could be congruent with either explanation.

References

Alaoui, L. and Penta, A. (2016). Endogenous depth of reasoning, *The Review of Economic Studies* **83**(4): 1297–1333.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). Measuring utility by a single-response sequential method., *Behavioral Science* **9**(3): 226–232.

Bors, D. A. and Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form, *Educational and Psychological Measurement* **58**(3): 382–398.

Chambers, C. P. and Hayashi, T. (2018). Reverse bayesianism: a comment, *American Economic Journal: Microeconomics* **10**(1): 315–24.

Charness, G., Karni, E. and Levin, D. (2007). Individual and group decision making under risk: An experimentalstudy of bayesian updating and violations of first-order stochastic dominace, *Journal of Risk and uncertainty* **35**(2): 129–148.

Charness, G. and Levin, D. (2005). How psychological framing affects economic market prices in the lab and field, *The American Economic Review* **95**(4): 1300–1309.

Chen, D. L., Schonger, M. and Wickens, C. (2016). otree - an open-source platform for laboratory, online and field experiments, *Journal of Behavioral and Experimental Finance* **9**: 88–97.

Dekel, E., Lipman, B. L. and Rustichini, A. (1998). Standard state-space models preclude unawareness, *Econometrica* **66**(1): 159–173.

Dietrich, F. (2018). Savage's theorem under changing awareness, *Journal of Economic Theory* **176**: 1–54.

Dominiac, A. and Tserenjigmid, G. (2017). Does growing awareness increase ambiguity?

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices., *Journal of Economic Behavior & Organization* **68**(1): 1–17.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments, *Experimental economics* **10**(2): 171–178.

Gennaioli, N. and Shleifer, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility.*, Princeton University Press.

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy* **124**(6): 1619–1676.

Grant, S., Meneghel, I. and Tourky, R. (2017). Learning under unawareness, *Available at SSRN 3113983* .

Grant, S. and Quiggin, J. (2013). Inductive reasoning about unawareness, *Economic Theory* **54**(3): 717–755.

Greiner, B. (2004). An online recruitment system for economic experiments., *Munich Personal RePec Archive* .

Halpern, J. Y. and Rêgo, L. C. (2008). Interactive unawareness revisited, *Games and Economic Behavior* **62**(1): 232–262.

Heifetz, A., Meier, M. and Schipper, B. C. (2006). Interactive unawareness, *Journal of economic theory* **130**(1): 78–94.

Holt, Charles A.and Smith, A. M. (2009). An update on bayesian updating, *Journal of Economic Behavior & Organization* **69**(2): 125–134.

Isoni, A., Graham, L. and Sugden, R. (2011). The willingness to pay—willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Comment., *American Economic Review* **101**(2): 991–1011.

Karni, E. (2009). A mechanism for eliciting probabilities, *Econometrica* **77**(2): 603–606.

Karni, E., Valenzuela-Stookey, Q. and Vierø, M.-L. (2020). Reverse bayesianism: A generalization., *The B.E. Journalof Theoretical Economics* **forthcoming**.

Karni, E. and Vierø, M.-L. (2013). "Reverse Bayesianism": A choice-based theory of growing awareness, *American Economic Review* **103**(7): 2790–2810.

Karni, E. and Vierø, M.-L. (2015). Probabilistic sophistication and reverse bayesianism, *Journal of Risk and Uncertainty* **50**(3): 189–208.

Karni, E. and Vierø, M.-L. (2017). Awareness of unawareness: a theory of decision making in the face of ignorance, *Journal of Economic Theory* **168**: 301–328.

Kochov, A. (2010). A model of limited foresight, *Technical report*, working paper, University of Rochester.

Machina, M. J. and Schmeidler, D. (1992). A more robust definition of subjective probability., *Econometrica* **60**(4): 745–780.

Machina, M. J. and Schmeidler, D. (1995). Bayes without bernoulli: Simple conditions for probabilistically sophisticated choice., *Journal of Economic Theory* **67**(1): 106–128.

Modica, S. and Rustichini, A. (1999). Unawareness and partitional information structures, *Games and Economic behavior* **27**(2): 265–298.

Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news, *American Economic Review* **102**(6): 2410–36.

Raven, J., Raven, J. C. and Court, J. (1998a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices.*, Oxford, UK: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J. C. and Court, J. (1998b). *Manual for Raven's progressive matrices and vocabulary scales. Section 1: General overview.*, Oxford, UK: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Schipper, B. C. (2013). Awareness-dependent subjective expected utility, *International Journal of Game Theory* **42**(3): 725–753.

Shackle, G. L. S. (1992). *Epistemics and economics: A critique of economic doctrines.*, Transation Publishers.

Sonnemann, U., Camerer, C. F., Fox, C. R. and Langer, T. (2013). How psychological framing affects economic market prices in the lab and field, *PNAS* **110**(29): 11779–11784.

Trautmann, S. T. and Schmidt, U. (2012). Pricing risk and ambiguity: The effect of perspective taking., *Quarterly Journal of Experimental Psychology* **65**(1): 195–205.

Trautmann, S. T. and van de Kuilen, G. (2015). Ambiguity attitudes., *in* G. Keren and G. Wu (eds), *The Wiley Blackwell Handbook of Judgment and Decision Making*, Blackwall, chapter 3, pp. 89–116.

Trautmann, S. T., Vieider, F. M. and Wakker, P. P. (2011). Preference reversals for ambiguity aversion., *Management Science* **57**(7): 1320–1333.

Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability, *Psychological Review* **101**(4): 547–567.

Viscusi, W. K. and Chesson, H. (1999). Hopes and fears: the conflicting effects of risk ambiguity., *Theory and decision* **47**(2): 157–184.

Appendices

Experiment 1 Instructions

Note: In red are comments that were not visible to participants. Any text highlighted in yellow relates to only the IS condition, while any text highlighted in green relates to only the PS condition

General Information

Welcome to the experiment.

Thank you for volunteering your time to participate in this experimental project. The purpose of this experiment is to study how people make decisions in a particular situation. The results of this experiment will have applications to behavioral economics and economics in general.

During this experiment, please follow the instructions very carefully. Please remain silent during the session. You will go through various stages where in some instances you will simply observe outcomes of draws from an urn and in other cases you will be asked to report your perceived likelihood of an event or how much an item is worth to you. These reports, which you will enter on your computer, will determine your eventual monetary reward from participating in this experiment.

You may have heard about experiments in which participants were deceived. This experiment ***does not*** *involve deception by the experimenters. That is, everything the experimenter tells you, and all on-screen instructions, are true and accurate.*

Initial Instructions

It is critical that you read through these instructions carefully as fully understanding them will allow you to substantially increase your eventual monetary payoff from this study ***, where you can earn from a minimum of €4.00 to a maximum of €26.00 depending on your decisions.***

During this study you will be asked to observe some outcomes of draws from an urn and will then be asked to report how likely certain 'events' are and how much some 'items' are worth to you. One of these choices will be randomly picked to determine your monetary payment for completing this study. Depending on your responses you stand to earn a substantial amount of money. Over the next few pages *we explain how your earnings will be determined. Please read this very carefully.*

Likelihoods of events – Reporting and Earnings

At different instances during this study, you will be asked to provide us with your perceived likelihood of an outcome of a random draw from an urn. The urn will contain prizes of varying monetary value. You will have an opportunity to observe multiple random draws out of this urn to gain a good understanding of the likelihood of different prizes.

After observing a sequence of random draws from the urn, you will be asked to report your beliefs about the likelihood that the prize drawn from the urn is of particular value. For example, we may ask you to report your perception of the likelihood that the prize drawn out of the urn has a value of €30.

You will be asked to report a number between 0 and 1. The closer to the true value your reported number is the greater would be your expected potential bonus from this decision.

Your best strategy is to estimate your own perceived likelihood and truthfully report that likelihood.

Suppose that we ask you to report the likelihood of drawing a ball with prize X. Your reported likelihood will be compared to a likelihood randomly generated by the computer. This ***randomly generated likelihood*** *will be a number between 0 and 1 and it will be completely unrelated to your reported likelihood. If your reported likelihood is greater than or the same as the randomly generated likelihood,* *you will be endowed with a lottery that pays you X with a probability equal to the actual proportion of prize X in the urn and pay you nothing otherwise. If, instead, your reported likelihood is lower than the randomly generated likelihood, you will be endowed with a lottery that pays you X with a probability equal to the randomly generated likelihood and pay you nothing otherwise. After these choices are made and revealed by the computer, the lottery in your (virtual) possession will be played and payments made according to the realization of the lottery.*

Example: Let's say we ask you to provide your perceived likelihood that a prize of value €30 will be randomly drawn from the urn. Let's further suppose that the true proportion of prize €30 is 0.5 in the urn. If your reported likelihood is 0.4 and the randomly generated likelihood is 0.3, you will receive the lottery that pays you €30 with probability 0.5 (true proportion of the prize) and pay you nothing otherwise. If your reported likelihood is 0.4 and the randomly generated likelihood is 0.6, you will receive the lottery that pays you €30 with probability 0.6 (the randomly generated likelihood) and pay you nothing otherwise.

Values of 'items' – Reporting and Earnings

At different instances during this study, you will be asked to provide us with the value at which you would be willing to sell an 'item'. This 'item' will be a lottery that would give you some monetary prizes with some probabilities. During the tasks that will follow, the prizes that will be possible to earn and their likelihoods will not be explicitly stated and so you would need to rely on what you expect.

For example, if you believe that there is equal likelihood of two prizes of value of €20 or €10, then the corresponding lottery would entail a payoff of €20 with probability 0.5 or a payoff of €10 with probability 0.5.

You will be given the lottery and it will be your task to provide us with a value which you would feel comfortable to sell us back this lottery. Your decision is essentially to provide us with the certain amount that you would be happy to receive instead of playing the lottery at the particular instance.

As you will see, your best strategy is to provide us with the minimum amount you would be willing to receive for selling us the lottery

Your named amount will be compared to a fixed amount. This fixed amount will be randomly generated by a computer and will be completely unrelated to your named amount:

- *If your named amount is less than or the same as the fixed amount, then you get to sell the lottery. But, here's the interesting part. You do not receive the amount you offered. Instead, we pay you the fixed amount, i.e. the randomly generated amount which is higher than or equal to your offer.*
- *If, on the contrary, your named amount is more than the fixed amount then you don't get to sell the lottery and will be paid according to the realization of the lottery.*

Example: if your named amount is €50 and the fixed amount is €60, you get to sell the lottery and receive the certain amount of €60. if your named amount is €50 and the fixed amount is €40, you do not get to sell the lottery and thus receive payment according to the realization of the lottery.

You should offer the minimum amount you would be willing to accept in exchange for the lottery you own. Your best strategy is to determine your personal value for the item and record that value as your offer. ***It will not be to your advantage to suggest more than this amount, and it will not be to your advantage to suggest less.*** *There is not necessarily a "correct" value. Personal values can differ from individual to individual.*

Example of best strategy for deciding valuation

The following example illustrates how you work out the minimum you are willing to accept for a lottery.

Imagine that I am a seller of a lottery "A". How do I know the minimum I'd be willing to sell lottery "A" for?

Start with 1 penny. Would I be willing to get 1 penny for the item? If NOT, then increase the amount to 2 pence. If I'm NOT willing to accept 2 pence, then increase further. I keep increasing until I come to an amount that makes me indifferent between keeping lottery "A" or getting a certain amount.

Example: Would I sell lottery "A" for €1.00? NO. So I need to consider higher amounts. Would I sell lottery "A" for €2.00? YES. Would I sell lottery "A" for €1.90? YES, Would I sell lottery "A" for €1.80, YES. Would I sell lottery "A" for €1.50? I don't care whether I end up with €1.50 or keep the lottery. Then that is the minimum I'd be willing to accept for lottery "A". I'll record that number on the computer.

The key to determining the minimum you'd be willing to accept is remembering that you will not necessarily *get* only *the amount you declare. Instead, if you receive anything, you will receive the fixed offer.*

Why is my best strategy to declare the minimum I'd be willing to accept? Let's go back to the example:

Say that I decide that the minimum I'd be willing to accept for lottery "A" is €1.50.

What happens if I declare more than €1.50? Say I declare €2.

If the fixed amount is, say, €1.90, then I don't sell the lottery. Had I declared €1.50, I would have received the amount €1.90 for a lottery that I think is worth €1.50. So I lose out.

What happens if I declare less that €1.50? Say I bid €1.00.

If the fixed offer is €1.20, then I have to accept €1.20 for a lottery that I really think is worth €1.50. I lose out.

Payment procedures

You will be asked to provide a value for lotteries and likelihood for events at different instances as we described in the previous pages. One out of all these decisions will be randomly chosen and payments will be made according to that decision and the realisation of the relevant lottery.

All prizes and lotteries that you will be asked to consider and make decisions about will be expressed in tokens. Each token corresponds to €0.05. Thus, a prize of 100 tokens will be equivalent to €5.

PART 1

You will now simply observe random draws of balls out of an urn. This urn contains a number of balls. Each ball represents a potential prize in terms of payment in tokens. You will observe 20 consecutive random draws with replacement from the urn. Please pay attention to the prizes that will appear and their frequencies. ***For IS conditions:*** *This urn contains* ***two and only two possible prizes.*** *Your earnings do not directly depend on the outcome of each random draw in this stage, but understanding the composition of the urn may considerably improve your future earnings.* ***For PS conditions:*** *At any point in the study new balls representing different tokens to what you have been observing so far may be added to this urn. Please click OK when you are ready to proceed.*

Belief Screen 1

We would now like to ask you some questions about the likelihoods of different prizes in the urn you have been observing.

Remember that your most profitable strategy is reporting truthfully your assessment of different likelihoods.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

What is your estimate of the likelihood that prize 80 is drawn from the urn? ______

What is your estimate of the likelihood that prize 190 is drawn from the urn? ______

WTA Screen 1

We would like to ask for your value of a lottery.

Recall that when answering questions about your value of a lottery, your named amount will be compared to a fixed amount. If your named amount is greater than or equal to the fixed amount then you do not sell the lottery and are thus paid according to the realization of the lottery. Otherwise, if your named amount is less than the fixed amount, you get to sell the lottery you have been endowed with and receive the fixed amount as a payment.

Remember that your most profitable strategy is reporting truthfully your valuations, you can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

> *Thinking about the different prizes and the composition of the urn you have just observed.*
>
> *What is the minimum amount you are willing to accept to sell the lottery that pays according to a draw from the urn? ______*

After giving a choice, the following appears on the screen

Thank you. Your choice has been recorded.

For IS conditions:

We will now draw another prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, the draw about to take place will be used to determine your payment.

You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

We make a draw out of the original urn and the participant making the draws notes down the prize drawn.

At this point we bring the new urn, draw one prize and say:

On PC Screen:

*This urn contains **only** the prize you are about to be shown. Please click OK to confirm you understand this.*

Now we drop the contents of the original urn into the new urn which together form the updated urn.

For PS conditions:

If the decision you just made is selected by the computer to be the payment relevant round, the draw about to take place from the urn in the front will be used to determine your payment.

At this point we bring the new urn, draw one prize and say:

On PC Screen:

This urn contains new prizes. The urn contains no prizes similar to what you have been observing as a result of random draws from the other urn. Please click OK to confirm you understand this.

Now we drop the contents of the original urn into the new urn which together form the updated urn.

We will now draw a prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, this draw will be used to determine your payment. You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

A random draw from the updated urn is made and the participant making the draws notes down the prize drawn.

Belief Screen 2

We would now like to ask again some questions about the likelihoods of different prizes in the urn you have been observing.

Remember that your most profitable strategy is reporting truthfully your assessment of different likelihoods.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

What is your estimate of the likelihood that prize 80 is drawn from the urn? ______

What is your estimate of the likelihood that prize 190 is drawn from the urn? ______

What is your estimate of the likelihood that prize 15/375 is drawn from the urn? ______

WTA Screen 2

Remember that your most profitable strategy is reporting truthfully your valuations.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

Again thinking about the urn in front of you.

What is the minimum amount you are willing to accept to sell the lottery that pays according to a random draw from the urn? ______

After giving a choice, this appears:

Thank you. Your choice has been recorded

We will now draw another prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, this draw will be used to determine your payment. You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

A random draw from the updated urn is made and the participant making the draws notes down the prize drawn.

PART 2

Lottery Choice Task

On your screen below you see a list of 6 lotteries with the prizes given in terms of tokens. You have to make a choice among these 6 lotteries. For each of the listed below lotteries, the chance for either of the two payoffs is equal. That is, for lottery 2 for example, you can win 24 tokens with 50% chance and 36 tokens with 50% chance. Your chosen lottery will be played out and you will be paid according to the realization of that lottery. As before, each token corresponds to €0.05 Thus, for a prize of 100 tokens the equivalent dollar amount will be €5.

Lottery	X	Y
1	28	28
2	24	36
3	20	44
4	16	52
5	12	60
6	2	70

PART 3

Short Raven Test implemented

PART 4

General Demographic Questionnaire

- How old are you? (years)
- What is your gender? (M/F/Other [Please describe if you wish]/Prefer not to disclose)
- What is your country of origin?
- What is your religion?
    - Budhist
    - Christian
    - Hindu
    - Jewish
    - Muslim
    - Sikh
    - No religion
    - Other [Please describe if you wish]
    - Prefer not to disclose
- What is your field of studies/major?
- What is your year of study?
- In high school, what was the highest possible grade? (E.g. A, 100, 20)
- What was your final grade in high school?
- In political matters, people talk of “the left” and “the right”. How would you place your views on this scale, generally speaking?

<u>Experiment 2 Instructions</u>

Note: In red are comments that were not visible to participants.

<u>General Instructions</u>

Thank you for participating in today's experiment.

If you have any questions during the experiment, please raise your hand. An experimenter will approach your table to answer your question in private.

You may have heard about experiments in which participants were deceived. This experiment does not involve deception by the experimenters. That is, everything the experimenter tells you, and all on-screen instructions, are true and accurate.

The experiment consists of 4 parts. For participating in this experiment you will earn £3 at the end of the experiment. In addition you can earn a bonus of £6 in each of the four parts, depending on your performance in the experiment and chance. After these four parts you will play a pattern game, in which you can earn additionally up to £2.

In the end follows a short demographic questionnaire.

<u>Sampling Boxes</u>

The experiment consists of 4 parts. In each part, you draw a random sample of (virtual) marbles from a (virtual) box containing exactly 100 colored marbles. Initially, you have no information about the contents of each box: you do not know which colors, or how many different colors, are in the box. The four parts and four boxes are independent of each other: different boxes are used for different parts.

In each part, you draw 30 marbles <u>with replacement</u> one after another from the box. You draw a marble by clicking the button "Draw" (or by pressing enter). Once clicked, the computer randomly draws a marble from the box. The result of a draw is shown on-screen with a marble of the color and the name of the color.

The sample draws are conducted with replacement. For example, if you drew a magenta marble (this color is not used in the actual experiment) from a box, this marble is placed back in the box for the next draw, such that the number of marbles of each color in the box stays the same as you sample. All marbles you have sampled (and their colors) are registered at the bottom of the screen.

Your payoff-relevant task

After each draw of a new marble, you will be asked to state your expectation about the contents of the box, that is, about the distribution of colors in the box. The more precise your prediction is, the higher will be your expected payoff from the experiment (details below). After each draw, you will be asked to separately indicate:

(i) Your expected number of marbles in the box for each color that you have already observed for the box, and

(ii) Your expected number of marbles of "any other colors" that you have not yet observed for the box, and that may or may not be in the box.

Example

Suppose you drew a magenta marble in your first draw and a teal marble (this color is also not used in the actual experiment) in the second draw. After the first draw you would be asked to guess how many magenta marbles are in the box, and how many marbles of any other color, not yet observed, are in the box. After the second draw you would be asked how many magenta marbles are in the box, how many teal marbles are in the box, and how many marbles of any other color, not yet observed, are in the box.

As the box contains exactly 100 marbles, your estimates of the number of marbles of the already observed colors and of any other colors you may think are in the box (but not yet observed) must add up to 100. Moreover, if you expect the number of marbles of other colors to be zero, you need to explicitly submit an estimate of zero (that is, not just leaving the entry field open).

After the 30th marble is drawn, you will enter your last prediction for this box. A new button "Continue to the next box" will allow you to continue to the next part, with a new box to sample.

Entering estimates in the program

After each draw, you can enter your estimates by typing them into the entry fields. You can also use the "fill previous estimate" buttons to pre-fill your previous round's estimates for each color. At any point before making the next draw, you can adjust the current estimates in the entry fields using "+" and "–" buttons next to the entry fields.

<u>Getting paid for good predictions</u>

You may earn a bonus of £6 for each part of the experiment. All of your answers provided for all four parts will affect your chances of receiving the bonus. If you want to maximize your expected earnings from this experiment, it is in your best interest to estimate the number of marbles for each box as accurately as possible, and report them truthfully after <u>each</u> draw. To determine whether you will win a bonus, you will draw a marble either from one of the boxes in the experiment (called *Estimate Box*), or from another, newly constructed one (called *New Box*). Importantly, your reported estimates will influence the construction of this new box.

If you report your estimates accurately and truthfully, this will be best for you in terms of your expected payment from the experiment. Below we will explain the <u>payment procedure</u>, and provide <u>the intuition</u> and an <u>example</u> why it is in your best interest to report your estimates as correctly as possible after each draw. You are invited to review these explanations. Please note that they are not necessary to understand the experiment and can be skipped without any harm if you are not interested. You can request a hard copy of these details at any point of the experiment in case of doubt.

Payment procedure (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

After you finished sampling from all four boxes, for <u>each</u> of the four parts you may earn a bonus of £6 as follows:

Estimate box: The computer randomly selects one of the 30 draw rounds, and then randomly selects one color estimate you made for this round (this is the selected color for this task). This can be an estimate for some color you have observed, or alternatively an estimate for the number of not yet observed colors at some point, that is, "any other color". Note than all of your estimates have the same chance to be randomly selected.

New box: Next, the computer constructs a new box of 100 marbles that contains only two colors, black or white. Every possible combination of black and white marbles (the number of white marbles = 100 – the number of black marbles) is equally likely.

Next, the computer compares the number of black marbles in the New Box with the estimate you made for the selected color in the experiment (or for "any other color").

- If your estimate for the selected color is <u>larger</u> than the number of black marbles in the New Box, you will draw one marble from the <u>Estimate Box</u>. If this marble is of the selected color, you will receive £6. If the marble is not of the selected color, you will receive £0.

- If your estimate is smaller than the number of black marbles in the New Box, you will draw one marble from the New Box. If this marble is black, you will receive £6. If the marble is white, you will receive £0.

Intuition (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

You will have the best chance to win the bonus of £6 for each part, by truthfully reporting your estimate. For example, if you think there are many magenta marbles in the Estimate Box, you will more likely make a draw from this box. This is because in your estimation the number of black marbles in the New Box will most likely be smaller than your estimate of magenta marbles for the Estimate Box.

If you think there are only few magenta marbles in the Estimate Box, you will more likely make a draw from the New Box. This is because the number of black marbles in the New Box will most likely be larger than your estimate of magenta marbles for the Estimate Box.

Thus, as long as you report your estimate for each color in each draw and each box accurately and truthfully, the mechanism makes sure that you get the box with the highest chance of winning the bonus.

Note that your winning chance in the case of making the payoff-relevant draw from the Estimate Box depends only on the true number of marbles of that color in the box. Similarly, in the case of making the payoff-relevant draw from New Box, the chance depends only on the number of black marbles in the box. Your estimate of colors for the boxes in the experiment is only relevant for determining the best boxes for you during the payment procedure. Thus, better estimates give you better chances to win.

Example (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

Example - Part 1

For part 1, the computer selected the round 16 draw. In this round you provided estimates of the number of magenta marbles, teal marbles, and the number of marbles of "any other color".

The computer further selected magenta as the payoff-relevant color estimate. Suppose your estimate of the number of magenta marbles in box 1 in round 16 was 42 marbles.

Suppose the computer randomly generated a New Box that contained 35 black and 65 white marbles. Because 35 black winning marbles in New Box is less than your estimate of 42 magenta winning marbles in Estimate Box 1, your bonus would be determined by Estimate Box 1. Note that your true chance to win the bonus of £6 would depend on the <u>true</u> number of magenta marbles in box 1. Suppose you drew a teal marble from Estimate Box 1. Your bonus for part 1 would be £0.

Example - Part 2

For part 2 box, the computer selected the round 2 draw. In this round you provided estimates of the number of magenta marbles, and the number of marbles of “any other color”. The computer further selected “any other color” as the payoff-relevant color estimate. Suppose your estimate for the number of “any other color” marbles in box 2 in round 2 was 50 marbles.

Suppose the computer randomly generated another New Box that contained 7 black and 93 white marbles. Because 7 black winning marbles in New Box is less than 50 winning marbles of “any other color” in Estimate Box 2, your bonus would be determined by a draw from Estimate Box 2. Note that your true chance to win the bonus of £6would depend on the true number of - marbles in box 2 that are not magenta. Suppose you drew a teal marble from box 2. Your bonus for part 2 would be £6.

Example - Part 3

For part 3 box, the computer selected the round 30 draw. In this round you provided estimates for the number of magenta marbles, the number of teal marbles, and the number of marbles of “any other color”. The computer further selected teal as the payoff-relevant color estimate. Suppose your estimate for the number of teal marbles in box 3 in round 30 was 20 marbles.

Suppose the computer randomly generated another New Box that contains 87 black and 13 white marbles. Because 87 black winning marbles in New Box is more than 20 twinning marbles of teal color in Estimate Box 3, your bonus would be determined by New Box 3. Suppose you drew a black marble from box 3. Your bonus for part 3 would be £6.

Example - Part 4

For part 4 box, the computer selected the round 7 draw. In this round you provided estimates for the number of magenta marbles, the number of teal marbles, and the number of marbles of “any other color”. The computer further selected “any other color” as the payoff-relevant color estimate. Suppose your estimate for the number of “any other color” marbles in box 4 in round 7 was 33 marbles.

The computer randomly generated another New Box that contains 27 black and 73 white marbles. Because 27 black winning marbles in New Box 4 is less than 33 winning marbles of “any other color” in Estimate Box 4, your bonus would be determined by a random draw from Estimate Box 4. Note that your true chance to win the bonus of £6would depend on the true number of marbles in box 4 that were neither magenta nor teal. Suppose you drew a teal marble from box 4. Your bonus for part 4 would be £0.

Pattern game

You will now play a pattern game, where you are asked to solve some puzzles

On the screen, you will see a set of abstract pictures with one of the pictures missing. You need to choose a picture from the choices below to complete the pattern.

You will have a total of 8 minutes to complete 12 such puzzles.

During these 8 minutes you will be able to move forwards and backwards and change your answers using the buttons and tabs on your screen.

At the end of the experiment, the computer will randomly draw two of the puzzles from the pattern game. Each puzzle has the same probability to be chosen. For each of the two puzzles that you solved correctly, you will earn an additional £1.

Once the 8 minutes have passed, the pattern game will be automatically submitted and you will proceed to the results. You can submit all your answers and wait for the others to finish once you reach the last puzzle by clicking on the button that will appear and be labelled "Finish and go to results".

A Additional analysis for Experiment 1

Figure 6: Distr. of the residual in original urn($\hat{p}^o_x = 1 - \hat{p}^o_{80} - \hat{p}^0_{190}$)

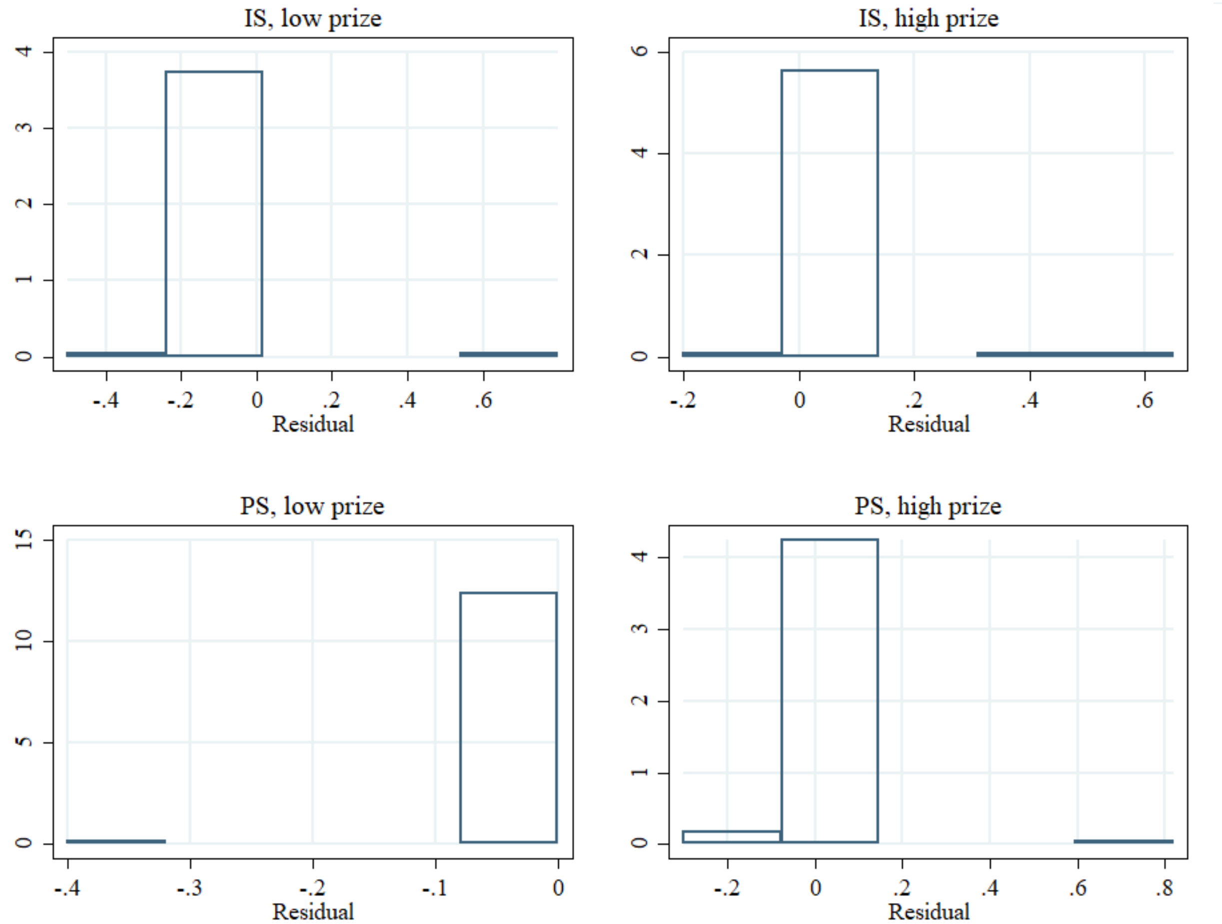


Figure 7: Distr. of the residual in updated urn ($\hat{p}^u_x = 1 - \hat{p}^u_{80} - \hat{p}^u_{190} - \hat{p}^u_s$)

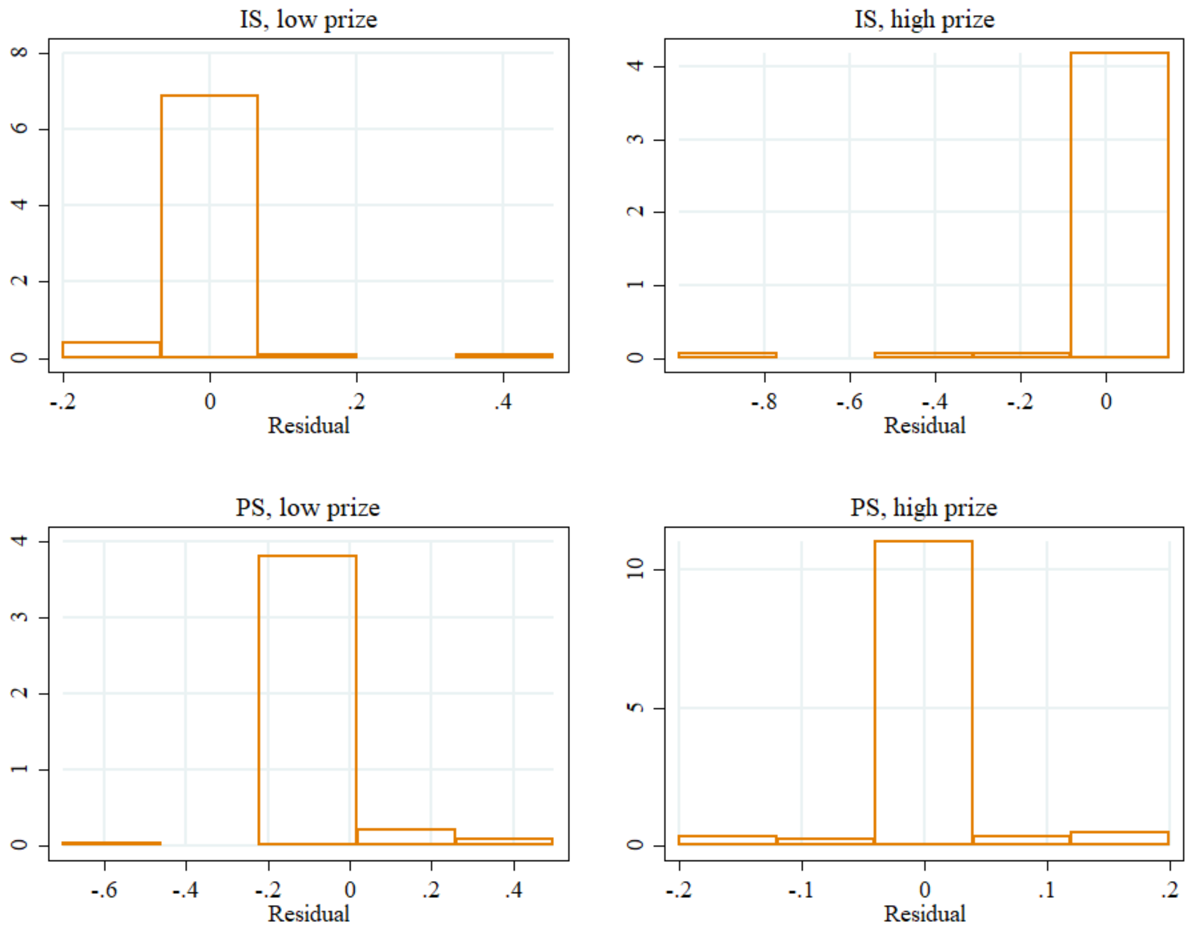


Table 12: Relation between ratio differences and Raven score; OLS regression.

Ratio differences	
Cognitive ability	-0.008
	(0.006)
Constant	0.234*
	(0.098)
Observations	343

* $p < 0.05$; ** $p < 0.01$
Standard errors in parentheses

Table 13: Relation between *WTA* and possible moderators, original urn; OLS regression.

WTA	All treatments	Low prize	High prize
PS	25.778**	27.530**	24.777**
	(6.438)	(10.285)	(8.240)
# prizes 80 observed	-1.958	-2.739	-0.724
	(1.585)	(2.191)	(2.650)
Cog Ability	1.482	2.799	0.243
	(1.511)	(2.604)	(1.827)
Age	-1.559*	-1.104	-2.161
	(0.776)	(1.125)	(1.102)
Female	-22.771**	-24.287*	-22.590**
	(6.487)	(10.515)	(8.301)
Econ	0.082	3.386	-2.151
	(7.388)	(11.465)	(9.824)
Constant	161.268**	147.290**	174.333**
	(26.602)	(40.866)	(37.721)
Observations	344	169	175

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

Table 14: Relation between *WTA* and possible moderators, updated urn; OLS regression.

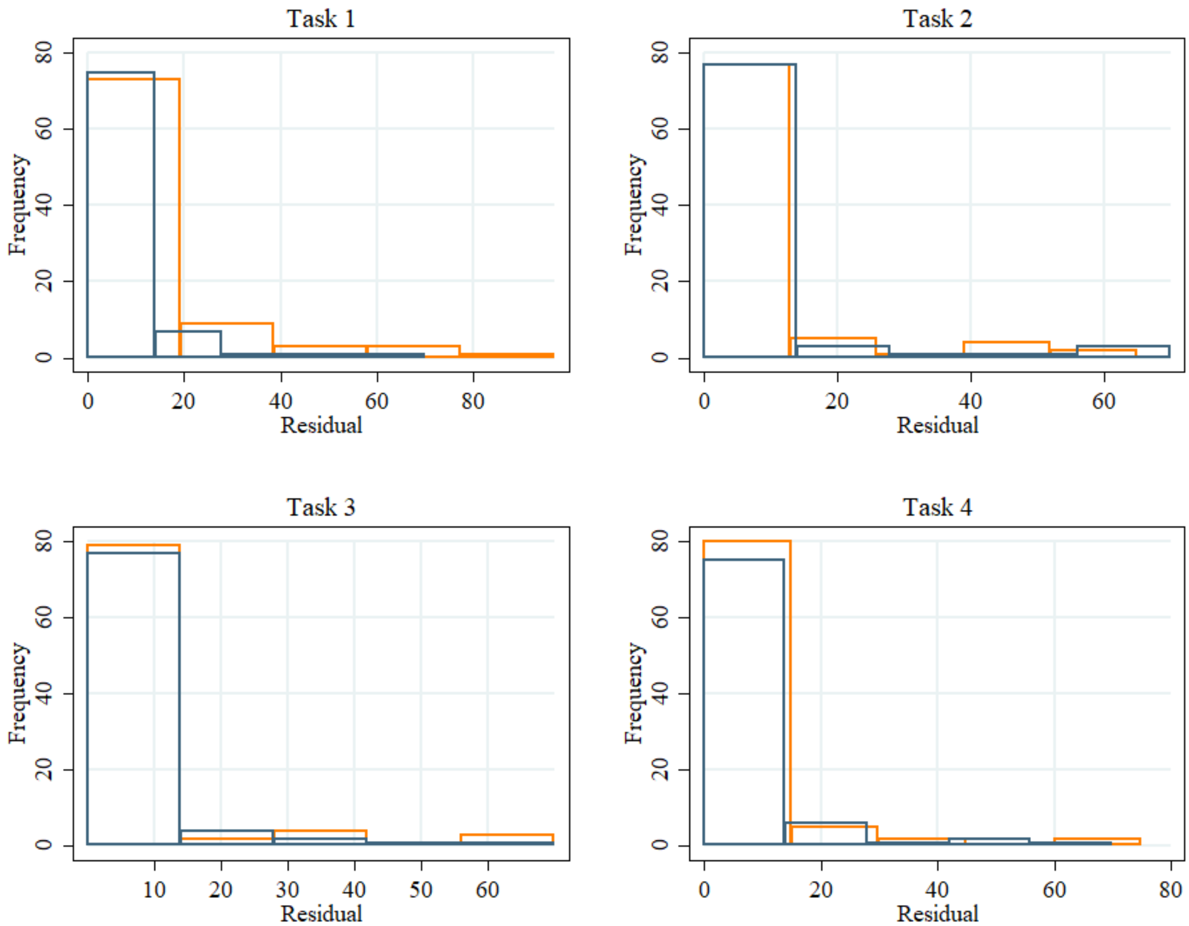
WTA	All treatments	Low prize	High prize
PS	18.326*	7.830	29.114**
	(7.740)	(7.206)	(10.796)
# prizes 80 observed	-2.604	-2.114	7.319*
	(1.906)	(1.535)	(3.473)
Cog Ability	2.196	4.238*	2.121
	(1.816)	(1.824)	(2.394)
Age	-0.865	0.802	-2.154
	(0.933)	(0.788)	(1.444)
Female	-15.170	-5.418	-32.600**
	(7.798)	(7.367)	(10.876)
Econ	0.009	14.545	1.081
	(8.882)	(8.032)	(12.872)
Constant	151.316**	52.566	142.509**
	(31.980)	(28.630)	(49.425)
Observations	344	169	175

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

B Additional analysis for Experiment 2

Figure 8: Residuals $\hat{p}_x$ after observing the last sample draw.



Orange boxes show residuals of *two colors* treatment, blue boxes show residuals of *four colors* treatment.

Table 15: Relation between the ratio differences and cognitive ability.

Ratio differences	Task 1	Task 2	Task 4
Cognitive ability	-0.010	-0.043	-0.042
	(0.024)	(0.038)	(0.030)
Num. draws	0.031**	0.045**	0.031**
	(0.009)	(0.014)	(0.011)
Num. colours observed	0.038	-0.114	-0.032
	(0.057)	(0.118)	(0.070)
Constant	1.025*	1.188	1.145
	(0.436)	(0.743)	(0.609)
Observations	1,119	160	630
Subjects			

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

Table 16: Difference in ratio changes, depending on if participants expected a surprise ($\hat{p}_x$>0).

	Didn't expect surprise	Expect surprise	z-statistic	p-value	p-value (corrected)
Task 1 ΔR^3	5	80	0.43	0.664	1.000
Task 1 $\Delta R^4_{i,1}$	5	79	-0.10	0.919	1.000
Task 1 $\Delta R^4_{i,2}$	5	79	1.28	0.202	1.000
Task 1 $\Delta R^4_{i,3}$	5	79	0.73	0.466	1.000
Task 2 ΔR^3	54	115	0.39	0.698	1.000
Task 4 ΔR^3	38	135	1.66	0.096	0.865
Task 4 ΔR^4_1	33	131	0.97	0.330	1.000
Task 4 ΔR^4_2	33	130	0.79	0.430	1.000
Task 4 ΔR^4_3	33	130	0.95	0.343	1.000

Wilcoxon rank-sum test, p-values corrected by Bonferroni-Holm procedure.

Table 17: Regression: Relation between samples and the residuals, fixed effects.

Size of the residuals	Panel OLS
Num. draws	-0.735**
	(0.046)
Num. colours observed	-8.495**
	(0.532)
Constant	52.451**
	(1.868)
Observations	20,880
Subjects	174

* $p < 0.05$; ** $p < 0.01$

Clustered standard errors in parentheses

Table 18: Regression: Relation between samples and the residuals after sample round 15, random effects.

Size of the residuals	Panel OLS
Num. draws	-0.349**
	(0.051)
Num. colours observed	0.852
	(0.795)
Cognitive ability	-2.299**
	(0.517)
Four colours first	-2.800
	(2.250)
Constant	35.562**
	(11.238)
Observations	9,900
Subjects	165

* $p < 0.05$; ** $p < 0.01$

Clustered standard errors in parentheses

Figure 9: Residual and the estimate of the new colour before and after observing the new colour, third colour.

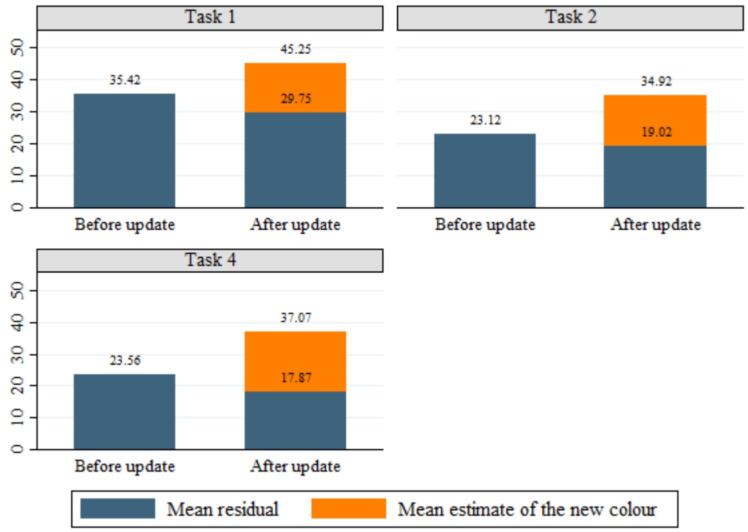


Figure 10: Residual and the estimate of the new colour before and after observing the new colour, fourth colour.