Retractions:
Learning from Information about Information

Duarte Gonçalves* Jonathan Libgober** Jack Willis***

*University College London
**University of Southern California
***Columbia University

February 14, 2023

Abstract. Toward understanding the persistence of widely discredited ideas, we study the effectiveness of retractions in correcting beliefs. Our experimental design identifies belief updating from retractions—the revoking of earlier information—and compares it to updating from equivalent new information. Retractions are ineffective: subjects update approximately one-third less from them versus both the retracted evidence and informationally-equivalent new evidence. Although we document several well-known biases in belief updating, our results require an explanation that treats retractions as intrinsically different. We find evidence for one such mechanism, while ruling out several others: retractions, information about information, are inherently more complex than direct information.

Keywords. Belief Updating, Retractions, Learning.
JEL Codes. D83, D91, C91.

Contact. duarte.goncalves@ucl.ac.uk, libgober@usc.edu, jack.willis@columbia.edu. An earlier version of this paper was circulated in June 2021 and then as an NBER working paper in November 2021 under the title "Learning versus Unlearning: An Experiment on Retractions." We benefited from helpful comments from many individuals, but we especially thank Dan Benjamin, Giorgio Coricelli, Cary Frydman, Terri Kneeland, Michael Thaler, Séverine Toussaert, Sevgi Yuksel, and seminar audiences at NYU, Caltech, Ohio State, USC, and the SWEET Conference. Jeremy Ward and Malavika Mani provided outstanding research assistance. Funding from IEPR is gratefully acknowledged.

1. INTRODUCTION

Retracted information often continues to influence beliefs, even once widely discredited. Baseless rumors, mistaken earnings announcements, fraudulent research findings, false claims of politicians; all tend to linger long after being debunked. Why is it so frequently easier to learn (incorrect) information than to subsequently "unlearn" it? While context-specific explanations for such retraction failures have been proposed across a variety of instances, their apparent ubiquity suggests a general underlying mechanism. Are retraction failures simply known failures of belief updating, such as confirmation bias, or is there something specific to the nature of retractions?

In this paper, we demonstrate that people underinfer from retractions relative to direct evidence and propose and find support for one mechanism, while ruling out many others: retractions—by their nature, information about information—are inherently more complex than direct evidence. First, we conduct a pre-registered experiment to quantify the degree to which retractions are less effective than direct evidence and to distinguish this finding from other belief updating biases. Second, we find that *complexity* plays a role in retraction failures by showing that updating from retractions is associated with common complexity indicators—longer response times and higher variability in responses. Third, we perform extensive robustness checks to assess and ultimately dismiss other possible explanations.

To this end, we rely on a canonical experimental design to identify and quantify retraction failures absent a variety of idiosyncratic confounds. Specifically, we develop a variation on the classic balls-and-urns experiment which is widely used to study limitations in information processing, for example in belief updating (Benjamin, 2019; Thaler, 2021; Ba, Bohren, and Imas, 2022), social learning (Anderson and Holt, 1997; Weizsäcker, 2010), and asset pricing (Halim, Riyanto, and Roy, 2019). Importantly, our design allows us to repeatedly provide retractions that are informationally equivalent to new observations, to subjects facing identical problems and sharing the same prior beliefs—aspects which are crucial to distinguish the effect of retractions from equivalent new information.

To study retractions, we modify the canonical balls-and-urns design as follows. Subjects are presented with draws of colored balls (blue or yellow) from a box with replacement, with one color being more likely depending on an underlying state. The box contains a "truth ball" which is either yellow or blue—the underlying state, over which we elicit subjects' beliefs—and "noise balls" in equal proportion. After presenting subjects with a series of such draws, in which they are told the color but not the truth/noise status of each ball, we then either present another such draw, or inform subjects whether a randomly chosen earlier ball draw was the truth ball or a noise ball. We refer to this latter event—when an earlier draw is disclosed to be a noise ball and thus uninformative of the underlying state—as a *retraction*. Eliciting beliefs on the underlying state, we test for retraction effectiveness by comparing beliefs following retractions to (a) beliefs without having observed the retracted observation in the first place, and to (b) beliefs following new draws with identical Bayes updates (in our setup, a draw of the opposite color to the retracted one).

Our first result identifies and quantifies retraction ineffectiveness as distinct from other belief updating biases. Subjects update less from retractions than from either (a) the retracted observation or (b) a new informationally-equivalent observation. Both results are robust across multiple variants of the experiment and hold regardless of details of the retraction; for example, whether information is confirmatory, or whether priors are more moderate or more extreme. The magnitude of retraction failures is large: beliefs update on average one-third less from retractions versus observations (see Section 5.3). In our theoretical analysis, we consider a general class of *quasi-Bayesian* models (which nests Bayesian updating—see Section 2.1 for a formal definition), and highlight that our results cannot be reconciled with any explanation that does not treat retractions as inherently different. This includes widely documented deviations from Bayesian updating such as base-rate neglect and confirmation bias, that our experiment replicates (see Benjamin (2019) for an authoritative survey).

Why are retractions less effective? We propose that retractions—information about information—are inherently more complex than direct information about the state, since they require additional contingent reasoning. Our proposed mechanism is motivated by two recent literatures: one showing that the need to consider more contingencies aggravates probabilistic errors in various domains (Ali, Mihm, Siga, and Tergiman, 2021; Esponda and Vespa, 2014, 2021; Martínez-Marquina, Niederle, and Vespa, 2019), and the other proposing that complexity considerations explain a number of well-documented behavioral biases (Oprea, 2020, 2022; Ba, Bohren, and Imas, 2022; Enke, Graeber, and Oprea, 2023).

To test this proposed mechanism, we take two approaches. First, we provide direct evidence that updating from retractions is harder, by considering process data and behavioral traits previously used in the literature to measure complexity of decision making. Second, we consider two instances where belief updating from or after retractions is *intuitively* more difficult, *show* that our behavioral markers confirm that the updating task is more complex in those instances, and then *test* whether updating strength following retractions changes as predicted by the complexity explanation.

The two behavioral markers of the complexity of an updating task we consider are decision times and the variability of belief reports, both of which have been used as measures of cognitive noise in past work—see e.g. Krajbich et al. (2012); Frydman and Jin (2022) for the former and Khaw et al. (2021) for the latter. Both proxies are larger when updating from retractions, as compared to equivalent new information, with updating taking 10% longer and leading to more than a one-third increase in variance. So, while previously we showed that updating is less effective from retractions than from equivalent direct evidence, these results suggest that it is also more cognitively demanding.

We examine how updating after retractions varies with complexity by leveraging natural variation in complexity provided by our design. Specifically, we compare belief updating (1) from retractions when earlier versus more recent observations are retracted, and (2) from new observations when there has versus has not already been a retraction. For (1), we argue that when the most recent observation is retracted, updating from a retraction is easier, since it requires simply returning to the previous belief. For (2), we suspect that if it is harder to update from retractions, it may also be harder to update from new observations *following* retractions, which we test by comparing updating when there was a previous retraction to when instead there was a previous informationally equivalent new observation. In both cases, our behavioral markers are aligned. Decision times are longer and belief variability larger: (1) when retractions do not refer to the most recent observation, and (2) when updating from new observations after observing a retraction. Correspondingly, subjects update less: (1) from retractions which do not refer to the most recent observation, and (2) from new observations if they have previously observed a retraction. Taken together, this evidence is consistent with our proposed mechanism and also supports our proxy measures for complexity.

We also consider alternative mechanisms in which retractions are treated inherently differently, and which could thus potentially explain our results. First, we document that the ineffectiveness of retractions is not caused by retractions exacerbating confirmation bias. In fact, they flip it: when updating from new observations, subjects slightly overinfer and do so more when observations confirm the prior—indicating confirmation bias—whereas when updating from retractions they underinfer and exhibit anti-confirmation bias. While this thus cannot explain the ineffectiveness of retractions, it again underlines how retractions are treated fundamentally differently than new observations, despite being informationally equivalent in our setting.

Second, we study whether our findings are driven by information being hard to disregard once it has been "acted" upon, as would be suggested by a cognitive dissonance explanation. If so, the ineffectiveness of retractions could be due to subjects having previously *used* the retracted observations to state their beliefs, before the retraction. To test this, we randomly assign subjects to a treatment arm of the experiment where we only elicit beliefs at the end of a sequence of observations and retractions, rather than after each draw. We fail to reject the hypothesis that retractions have the same effect on belief updating as in our baseline setup.

Finally, we test whether our results simply reflect some limited understanding of the subjects regarding the data generating process. We are able to rule out misinterpreting that the draws are made with replacement. We also consider removing subjects who are 'noisy' or prone to mistakes (e.g. updating in the wrong direction), or who did not correctly answer at first try unincentivized comprehension questions. A theme that emerges is that our results are stronger when restricting to subjects who appear to have understood the task better.[1] We conclude that our findings are not an artifact of some consistent misinterpretation of the design. Although we are not powered for a fully-fledged within-subject analysis, inspection of individual heterogeneity in our results indicates that the ineffectiveness of retractions compared with new observations is a general phenomenon in our sample.

Related to retraction failures is the idea that discredited information may retain a residual impact, known in psychology as the *continued influence effect* (see e.g. Johnson and Seifert, 1994; Ecker et al., 2022). Lewandowsky et al. (2012) surveys this literature and highlights several explanations; none appear capable of explaining our results, given our design. To the best of our knowledge, all past experiments on retractions involve information that is (at least partially) subjective, and leave open the possibility that subjects are interpreting them correctly within their subjective worldview. Further, our theoretical framework highlights that in some implementations of retractions, Bayesian updating should *not* entail simply "deleting" the retracted evidence. Moreover, since equivalent new information is not presented in these experiments, they do not separately identify retraction failures from other well-known biases, such as confirmation bias. Our contribution relative to this literature is to focus on a setting with objective probability

[1] This is perhaps unsurprising, since to find any effect requires subjects to act differently for retractions; if subjects answered randomly or always answer 50-50 we would not document any difference. In contrast, it is worth emphasizing that most of our sample did very well on unincentivized comprehension questions, confirming our assertion that our design achieved its desired simplicity despite also containing sufficient richness to define retractions and speak to mechanisms.

assessments, in a "context-free" setting where retractions can be compared to equivalent direct evidence. Such updating from retractions—or information about information in general—appears relatively unexplored in economics.

Our results are both of practical value and theoretical interest. We designed the experiment to provide the first evidence that retraction failures could be attributed (at least in part) to errors in information processing. From a theoretical standpoint, our findings motivate the development of theoretical models of costly information processing that treat information about information differently from direct information, even when their informational content is the same. From a practical standpoint, our analysis provide guidelines regarding how and when individuals can be expected to update beliefs with information about information, of potential relevance for campaigns targeting misinformation.[2] The fact that retraction failures arise from an information processing error suggest limits to the "this time is different" logic policymakers may adopt—it is in general unreasonable to expect a retraction to result simply in the "deletion" of a piece of information. We suspect that in many real-world cases, appreciating the inability to correct beliefs with retractions ex-post would have changed the calculus regarding decisions to disseminate information ex-ante.[3]

After presenting our theoretical framework in Section 2, Section 3 lists our main hypotheses and Section 4 presents our experimental design and implementation. Our analysis follows in three parts: Section 5 documents retraction failures, Section 6 examines mechanisms (both our main proposal and others), and Section 7 discusses robustness. Section 8 concludes.

2. FRAMEWORK

This section presents formal definitions and includes our main framework. We establish two key points which play a role in our subsequent analysis. First, we highlight the equivalences between retractions and new draws in terms of (quasi-)Bayesian updates. Second, we illustrate how complexity, response times, belief variances, and updating should relate to one another, which we use in analyzing our mechanism.

[2] In this sense, our paper is part of a sizable literature which, while motivated by anecdotal or domain-specific evidence of biases, utilizes basic belief updating tasks to highlight a relevant theoretical mechanism; see e.g. Oprea and Yuksel (2022), Esponda, Oprea, and Yuksel (2022), Hartzmark, Hirshman, and Imas (2021), or Agranov et al. (2022).

[3] We do not speak to issues of how these biases interplay with information *preferences*, although this might influence some of these decisions in practice; see Masatlioglu, Orhun, and Raymond (2021), Gul, Natenzon, and Pesendorfer (2021), Ambuehl and Li (2018) or Charness, Oprea, and Yuksel (2021) for papers studying this element.

2.1. Updating Beliefs from New Observations

We consider a decisionmaker who forms beliefs over a state θ, which takes one of two values with equal probability, say $\theta \in \{-1, 1\}$. At time t, the decisionmaker observes $s_t \in \{-1, 1\}$, informative of the state θ and independent conditional on θ. Throughout, we use the term "signal" as a generic term for information, and "observation" or "draw" for signals s_t that provide *direct* information about the state. We use $P(\cdot)$ to denote *objective* probabilities associated with the data generating process, and $b(\theta \mid \cdot)$ to denote the decisionmaker's *subjective* beliefs about the state.

Each observation s_t can either be *true*, in which case $s_t = \theta$, or *noise*, in which case it is given by an independent ϵ_t. Denoting the former event by $\{n_t = 0\}$ and the latter by $\{n_t = 1\}$,

$$s_t = (1 - n_t) \cdot \theta + n_t \cdot \epsilon_t, \tag{1}$$

where $n_t \in \{0, 1\}$ and n_t, ϵ_t and θ are independent. For simplicity, we write $S_t = \{s_1, \ldots, s_t\}$.

For a Bayesian, $b(\theta \mid S_t) = P(\theta \mid S_t)$. Past work has routinely rejected this hypothesis. One way to test for deviations from Bayesian updating (see Benjamin, 2019) is to note that log-odds updates are constant when observations are identically distributed; that is, if $K(s_{t+1}) = \log\left(P(s_{t+1}|\theta)/P(s_{t+1}| - \theta)\right)$; then for a Bayesian decisionmaker the following equation

$$\log\left(\frac{b(\theta \mid S_{t+1})}{b(-\theta \mid S_{t+1})}\right) = \alpha \log\left(\frac{b(\theta \mid S_t)}{b(-\theta \mid S_t)}\right) + \beta K(s_{t+1}), \tag{2}$$

should hold for $\alpha = 1$ and $\beta = 1$. Base rate neglect, for instance, corresponds to the hypothesis that $\alpha < 1$; underinference corresponds to the hypothesis that $\beta < 1$.

Since at least Kahneman and Tversky (1979), a common alternative is to instead assume there is a strictly increasing probability weighting function f such that

$$b(\theta \mid S_t) = f(P(\theta \mid S_t)).$$

Even if $\alpha \neq 1$ or $\beta \neq 1$, as long as f is strictly increasing, it is invertible, so $f^{-1}(b(\theta \mid \cdot)) = P(\theta \mid \cdot)$. It then follows that $b(\theta \mid \cdot)$ is given by the following identity:

$$\log\left(\frac{f^{-1}(b(\theta \mid S_{t+1}))}{f^{-1}(b(-\theta \mid S_{t+1}))}\right) = \log\left(\frac{f^{-1}(b(\theta \mid S_t))}{f^{-1}(b(-\theta \mid S_t))}\right) + K(s_{t+1}), \tag{3}$$

As long as some f exists such that $b(\theta \mid \cdot) = f(P(\theta \mid \cdot))$, one could recover f by using (3). Inspired by Cripps's (2021) axiomatic work, we call such a decisionmaker "quasi-Bayesian:'

Definition 1. *We say that a decisionmaker is a "quasi-Bayesian" if there exists a strictly increasing f such that $b(\theta \mid s)$ can be derived from $b(\theta)$ by (i) computing $f^{-1}(b(\theta))$, (ii) determining $f^{-1}(b(\theta \mid s))$ using (3), and (iii) composing the result with f to obtain $b(\theta \mid s)$.*

Note that, to accommodate some forms of confirmation bias, it may be necessary to allow the function f to depend on the initial belief $b(\theta)$ subjects update from; we strive to be as agnostic as possible and our comparisons will hold across a number of possible assumptions.

Updating rules satisfying this requirement are commonly used in experimental work (e.g. Angrisani et al., 2021). Among possible microfoundations for such distortion is the hypothesis that the agent faces some cognitive imprecision, as posited by models of cognitive uncertainty, efficient coding, and sequential sampling.[4] In the spirit of the models in this literature, consider a situation in which our decisionmaker faces uncertainty about how to interpret the likelihood of evidence s_{t+1} and update beliefs. Suppose the decisionmaker's prior is given by $K(s_t) \sim \mathcal{N}(0, \sigma^2)$, and that by deliberating the decisionmaker obtains n estimates $K(s_{t+1}) + \sigma_\zeta \cdot \zeta_i$ with additive Gaussian noise $\zeta_i \sim \mathcal{N}(0, 1)$, until becoming sufficiently certain about $K(s_{t+1})$.[5] This yields posterior log-odds updates similar to the above:

$$\log\left(\frac{b(\theta \mid S_{t+1})}{b(-\theta \mid S_{t+1})}\right) = \log\left(\frac{b(\theta \mid S_t)}{b(-\theta \mid S_t)}\right) + \beta K(s_{t+1}) + \beta \frac{\sigma_\zeta}{\sqrt{n}} \zeta_i,$$

with $\beta = \dfrac{\sigma^2}{\sigma_\zeta^2/n + \sigma^2}$. It is then possible to characterize the agent's expected posterior belief by a probability weighting function such that $b(\theta \mid \cdot) = f(P(\theta \mid \cdot))$, with underinference being intrinsically associated to the agent's cognitive uncertainty.

As a prelude to our analysis of mechanisms later, we highlight some testable predictions that emerge. Suppose one finds that β decreases in the above equation. According to this model, this could be generated by (a) a decrease in n, (b) an increase in σ_ζ, or (c) a decrease in σ. A standard approach in the literature associates n with response time, the idea being that the decisionmaker obtains one such signal per unit of time spent deliberating (see footnote 4). This rationalizes the

[4] While distinct, the literatures are closely related. Efficient coding (Wei and Stocker, 2015) and cognitive uncertainty models have been increasingly popular in economics; e.g. Khaw et al. (2021), Frydman and Jin (2022), Enke and Graeber (2020), and Thaler (2021). Models of sequential sampling provide a relationship between cognitive uncertainty and time through evidence accumulation (Krajbich et al., 2010; Bhui and Gershman, 2018). See Ratcliff et al. (2016) for a survey of sequential sampling models in psychology and neuroscience, and Fudenberg et al. (2018), Alós-Ferrer et al. (2018), and Gonçalves (2022) for recent applications in economics.

[5] Since the purpose of this formalism is to fix ideas about existing concepts and relate cognitive uncertainty with "quasi-Bayesian" updating, we purposefully leave the threshold for 'sufficient certainty' as exogenous.

general finding that decisionmakers take more time on simple tasks when these tasks become less immediately apparent.[6,7]

2.2. Updating Beliefs from Retractions

We now turn to updating *from retractions*, highlighting subtleties that emerge in pursuit of showing that retractions are treated fundamentally differently from other information. Formally:

Definition 2. *Consider the data-generating process described in (1). A retraction at time t consists in informing the decisionmaker that $n_\tau = 1$, thus implying the observation s_τ was noise.*

For reasons described below, we focus on the following:

Definition 3. *A* verifying retraction *is a retraction in which $\{\tau = \ell\}$ is independent from other observations' truth value, $\ell \leq t$.*

In our experiment, this is implemented by selecting τ uniformly at random from $\{1, ..., t\}$ and subsequently revealing n_τ to the decisionmaker; that is, whether this observation is noise or not.[8] Note that a Bayesian decisionmaker should be able to follow Bayes rule and update beliefs following retractions without any ambiguity.[9]

The following result relates the following three quantities, our main comparisons in the paper:

(1) $b(\theta|S_t, n_\tau = 1)$, the decisionmaker's belief after observing the retraction $n_\tau = 1$;
(2) $b(\theta|S_t \setminus s_\tau)$, the decisionmaker's belief had the retracted observation s_τ never been observed.
(3) $b(\theta|S_t \cup s_{t+1})$, the decisionmaker's belief after observing a new observation s_{t+1} instead of the retraction

Proposition 1. *Suppose retractions are verifying. For any quasi-Bayesian updating rule with a fixed (initial-belief independent) f, as described in Definition 1, (1) and (2) are identical. Moreover, given any (possibly initial-belief dependent) quasi-Bayesian updating rule f, (1) and (3) are identical if and only if its loglikelihood is negative of the retracted observation, $K(s_{t+1}) = -K(s_\tau)$.*

[6] Early versions of these results can be found in, for instance, Banks et al. (1976), Buckley and Gillman (1974), or Ratcliff (1978); to our knowledge, scientific consensus accepts the basic finding.

[7] We note that to conclude that updating is harder at a given history, it is necessary to compare jointly how β (responsiveness to beliefs), n (reaction time), and $\frac{\sigma_\zeta}{\sqrt{n}}$ (belief variance). At a given history—so that σ is fixed—in order to rationalize (a) *increase* in n given (b) a decrease in β, it must also be the case that (c) $\sigma_\zeta/\sqrt{n}$ increases.

[8] Note that this implies that when $n_\tau = 0$, the decisionmaker learns their past information was actually true and, in the current setting, this would result in degenerate Bayesian posterior beliefs.

[9] This lack of ambiguity distinguishes our experiment from Liang (2020), Shishkin and Ortoleva (2021), and Epstein and Halevy (2020).

The proof of this proposition essentially follows from a careful application of Bayes rule and observing that quasi-Bayesian updating rules still satisfy this identity under the transformation f^{-1}. An identical argument could be used to introduce additional history dependence into the updating rule; our identification strategy below would remain valid. But while one may wish to entertain a variety of models to accommodate a plethora of biases, any differences between (1) and (2) or (3) in our experimental setup will require retractions to be treated as intrinsically different.

Our hypothesis will be that, when the updating task is harder, subjects experience higher cognitive uncertainty about how to update beliefs, resulting beliefs being less affected by information on average (i.e. smaller expected changes in beliefs $b(\theta|S_{t+1}) - b(\theta|S_t)$), and higher variance in belief updates.[10] This association between noise in belief updating, decision time, and task difficulty has been noted before and is in line with the literatures on cognitive uncertainty, efficient coding, and sequential sampling models, and their empirical findings (see, e.g. Frydman and Jin, 2022; Frydman and Nunnari, 2022). We return to these predictions in the following section, when discussing our hypotheses.

Our focus on verifying retractions makes it simplest for subjects to update correctly. This stands in contrast to other setups, in which subjects need to account for how signals provided by an information structure are *restricted*, a feature that Miller and Sanjurjo (2019) argue is responsible for mistakes in probabilistic reasoning.[11] While a retraction reveals that a give observation does *not* correspond to the state, our implementation of retractions is *unrestricted*, thus eliminating the concern that failure to update from retractions would be a mere expression of this pervasive limitation. Crucially, if information about past evidence were disclosed only when the evidence is found to be uninformative of the state, retractions would give more credence to *non-retracted* evidence.[12] Instead, with verifying retractions, an observation is selected independently of its truth value and a retraction asserts the noise value of the retracted observation, but provides no additional information about the truth value of *other* observations. Although these issues

[10] For instance, taking the number of estimates n positively related to time and as increasing in prior variance σ^2 and estimate variance σ_ζ^2, there are conditions under which our illustrative model exhibits these traits.

[11] Perhaps the most prominent example is the *Monty Hall Problem*, where a subject selects one of three doors, only one of which hides a prize. After making a choice, one of the *un*selected doors that does *not* hide the prize is revealed. The subject is then offered to switch their choice. Since only unselected doors *without a prize* can be revealed, the other unselected door is then more likely to hide a prize and it is optimal to switch. Friedman (1998) shows that subjects err with striking consistency, choosing often to keep their choices.

[12] For non-verifying retractions like this, the additional restrictions on which observations are retracted can be meaningful, similarly to what occurs in the Monty Hall problem. Indeed, Proposition 1 is no longer true if retractions provide information related to how other evidence was generated.

are certainly relevant in a number of circumstances, verifying retractions seem to be the natural starting point for our analysis.[13] We then deliberately preclude this phenomenon and focus on verifying retractions to make updating not only as simple as possible, but especially to make it equivalent to deleting retracted evidence and nothing more.

3. HYPOTHESES

The goal of the paper is to study updating from retractions and to compare it to updating from new observations. Correspondingly, our first hypothesis has two parts: part (a) concerns the failure of retractions to correct beliefs while part (b) compares retractions to equivalent new information:

Hypothesis 1 (Retractions are Ineffective)**.** *Subjects (a) fail to fully internalize retractions, and (b) treat retractions as less informative than an otherwise equivalent piece of new information.*

We emphasize that the use of the term "retractions" in this hypothesis reflects the meaning in Definition 2, with "otherwise equivalent" reflecting the last case of Proposition 1. Thus, this hypothesis conjectures that retraction failures can emerge as a (specific) departure from Bayesian updating without context-specific elements.

Note, importantly, that while (a) and (b) both reflect retractions being less effective, and that one conclusion may be *suggestive* of the other, they are ultimately distinct. In principle, both new observations and retractions could be treated as equivalent and less informative than an earlier observation, leading to (a) without (b)—retraction failures could be driven by a feature of learning common to both retractions and new information. Conversely, new observations and retractions could be treated differently, but with retractions being internalized fully and a distinct departure from Bayesian updating yielding overreactions to new observations, leading to (b) without (a).

Our next hypotheses concern our proposed explanation for why retractions are less effective than new observations: retractions are harder to process. That is, we conjecture that an extra layer of complexity is introduced for retractions, as subjects must consider what a retraction implies about past information; since new observations (without retractions) are exchangeable, this step is not required when learning from new observations alone. To test it, in line with our

[13] In ongoing research we examine a version of this experiment using targeted (i.e., non-verifying) retractions; the results are largely consistent, although direct comparisons between the two are unwarranted, as then it is not in general true that $P(\theta \mid S_t, n_\tau = 1) = P(\theta \mid S_t \setminus s_\tau) = P(\theta \mid S_t \cup -s_\tau)$. These results are available from the authors upon request.

discussion in the previous section, we consider two proxies for increased cognitive complexity: longer decision times and greater belief report variance. We conjecture that both will reflect the additional complexity inherent to this kind of conditional reasoning:

Hypothesis 2 (Retractions are Harder)**.** *Processing retractions is more difficult than processing new observations, resulting in longer decision time and greater belief variance.*

We also exploit the dynamic information arrival in our experimental design to test whether intuitively 'harder' retractions are less effective and similarly increase decision time and belief variance. In particular, when earlier observations are retracted, there is a layer of added complexity in belief updating: disregarding retracted observations entails forming beliefs about a dataset not previously observed—a difficulty potentially aggravated in the presence of base-rate neglect. In contrast, retracting the most recent observation only requires returning to the belief held prior to that observation, $b(\theta \mid S_t, \text{Retraction of } s_t) = b(\theta \mid S_{t-1})$.

Insofar as retraction failure is tied to their inherent complexity, retractions targeting more (less) recent information would be more (less) effective. This motivates the following hypothesis:

Hypothesis 3 (Harder Retractions are Harder)**.** *Retractions of less recent observations are (a) less effective, and (b) result in longer decision time and greater belief variance.*

If a retraction is harder to process, then it is plausible that it is also harder to update from new observations following a retraction. This would constitute another expression of our proposed mechanism, which we articulate as a related hypothesis:

Hypothesis 4 (Updating after Retractions)**.** *Processing new observations after retractions is more difficult, resulting in (a) subjects updating less from new observations, and in (b) longer decision time and greater belief variance.*

We conclude by considering alternative explanations for retraction ineffectiveness. Our design deliberately shuts off common explanations for retraction ineffectiveness—e.g. imperfect memory, motivated reasoning, complex narratives, reliability of the source of retractions—and our theoretical framework shows that retractions being less effective *requires* them to be treated differently from new observations. However, being treated differently does not necessarily imply that the same biases in belief updating are not present. We then consider if retraction ineffectiveness is simply an expression of well-known biases:

Hypothesis 5 (Similar Updating Biases)**.** *The same biases in belief updating from new observations are present in updating from retractions.*

Lastly, we consider a further alternative explanation for retraction ineffectiveness: that retractions are less effective simply because it is difficult to disregard evidence that has been acted on or engaged with. In contrast to an inherent greater complexity of retractions, this alternative mechanism relies on a form of cognitive dissonance or an endowment effect applied to information, suggesting the following hypothesis:

Hypothesis 6 (Retracting Used Evidence)**.** *Retractions are ineffective only when individuals have acted on the retracted observations.*

Taken together, our hypotheses posit a diminished effectiveness of retractions, address several implications of our proposed mechanism—that retractions are harder to process—and consider two alternative mechanisms that could underlie this outcome.

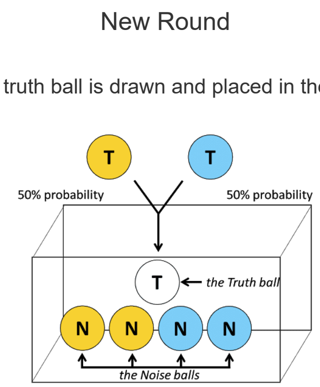
4. EXPERIMENTAL DESIGN

In this section, we describe the overall experimental design—visually summarized with screenshots from the experimental interface in Figure 1—and then provide details on the experimental interface and protocols. As our goal is to directly test the theoretical framework in Section 2, the basic data-generating process matches the theoretical framework presented there. Subjects were provided full information on how observations would be drawn.

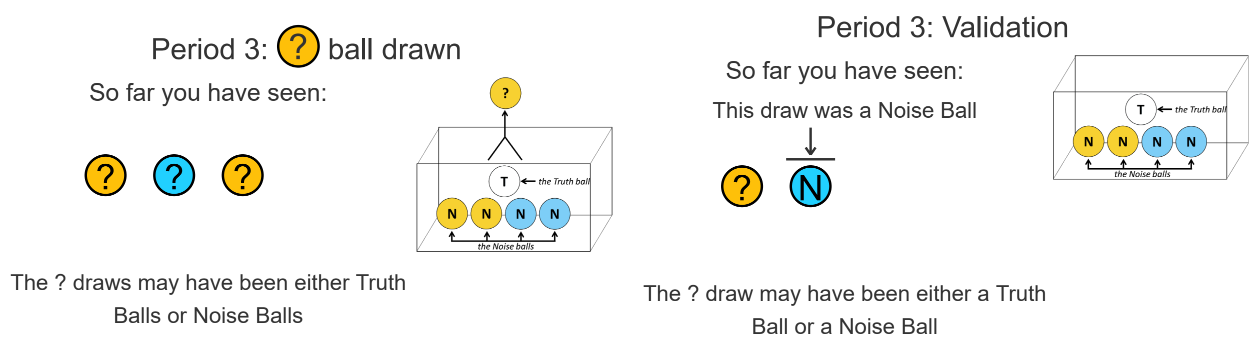
4.1. Basic Design

We first describe one round of the basic experimental design. Each *round* of the experiment has up to four *periods*, with beliefs elicited at the end of each period. Each subject plays a total of 32 rounds, and no feedback on performance is provided until the end of the experiment, when performance-based payouts are made. In each round, the sequence of events is as follows:

1. At the start of the round, a *truth ball* (referring to the state θ) is chosen at random to be either yellow or blue, with equal probability. The truth ball is then placed into the box with four *noise balls*, two yellow and two blue (corresponding to $P(n_t = 1) = 1/5$ and $P(\epsilon_t = 1) = 1/2$ in the information arrival process described in Section 2).



(a) *Determining the state.* At the beginning of each round, a truth ball was selected at random, with equal probability of being yellow or blue, and placed into a box with four noise balls, two yellow and two blue.



(b) *Ball draws and retractions.* Rounds consisted of (up to) four periods, each of which consisted of either a new draw or a verification, followed by the elicitation of subjects' beliefs over the color of the truth ball. For a new draw (left), a ball was drawn from the box (with replacement), and subjects were told its color but not whether it was the truth ball or a noise ball. For a verification (right), an earlier draw was chosen at random, and subjects were told whether that ball was a noise ball (a retraction) or the truth ball. If it was the truth ball, the round ended. The history of the round was displayed throughout.

Figure 1: Summary of Experimental Visuals

2. In periods one and two, subjects obtain a *new observation*: a draw from the box, with replacement. They are told the ball's color but not whether it is the truth ball or a noise ball.
3. In periods three and four, and independently across periods, subjects either obtain a new observation (as above), with probability 1/2, or they observe a verification of an earlier observation from the same round, with complementary probability. Under a verification, one of the prior draws is chosen at random and it is revealed whether it was a noise ball—a *retraction*—or the truth ball. If the draw is revealed to have been the truth ball, the round ends, as at that point the state (the color of the truth ball) is fully revealed.

Additionally, at the end of each period—that is, after each new signal (observation or retraction)—

subjects report their belief regarding the probability that the truth ball is blue vs. yellow. These reports are incentivized, as detailed later in the section.

To summarize, updating from retractions in our setup is made as simple as possible.

- The prior about the state and the noise are both symmetric ($P(\theta = 1) = P(\epsilon_t = 1) = 1/2$).
- Observations are independent and identically distributed conditional on the state and the log-likelihood of their realizations is symmetric around zero ($K(s_t) = -K(-s_t)$). This is necessary and sufficient for retracting s_τ to be equivalent both to deleting the retracted observation *and* receiving a new opposite observation $s_{t+1} = -s_\tau$ (Proposition 1).
- The details of the data-generating process are graphically described in an intuitive manner; these and the history of signals (observations and retractions) are always visible to subjects.

As explained in detail in Section 5.1, this simple design allows us to identify the effect of retractions on belief updating (Hypothesis 1). Since the equivalence in Proposition 1 holds for a broad class of belief updating rules—including Bayesian updating and generalizations common in the literature—and does not depend on nor does it require any information on past observations, variation across different histories will also allow us to test when retractions are more or less ineffective (Hypotheses 3, 5) and if retractions affect subsequent updating (Hypothesis 4).

4.2. Single-Elicitation Treatment

Our experiment features a between-subject treatment. At the start of the experiment, subjects are randomly allocated to one of two treatments. With 1/2 probability they are allocated to the *baseline treatment*, as described above. With 1/2 probability, they are allocated to the *single-elicitation treatment*, whose purpose is to test if requiring subjects to report their beliefs in *every* period—and hence to act on draws before they are retracted—affects the efficacy of retractions (Hypothesis 6).

In the single-elicitation treatment, the sequence of events is the same as in the baseline treatment, except for two differences: (1) beliefs are only elicited at the end of each round, rather than each period; (2) with probability 1/3, the round ends in period two; with probability 2/3, the round ends in period three. The design ensures that while we do not observe the *entire* belief path, we are nevertheless able to form estimates for beliefs after two draws, as well as beliefs after three draws when the third draw is either a retraction or a new observation.

4.3. Implementation Details

Experimental Interface. A summary of the explanatory visuals shown to subjects is given in Figure 1 and the full instructions of the experiment can be found in Online Appendix C. Beliefs were reported using a slider, which displayed both the probability they assign to the truth ball being either yellow or blue. After the instructions, subjects were given two rounds of unincentivized "practice" to familiarize themselves with the interface.

Subject Pool and Comprehension Checks. The experiment was run on Amazon Mechanical Turk (henceforth MTurk) on June 16-18, 2020. In order to ensure adequate statistical power, we targeted 200 subjects per treatment group. We recruited a total of 415 subjects, 211 subjects for our baseline setup and 204 for the single elicitation treatment. We took several steps to ensure that our subject pool was of high quality—these are described in greater detail in Section 7.

Payments. We incentivized subjects to report their beliefs truthfully using a binarized scoring rule (Hossain and Okui, 2013; Mobius et al., 2022). By reporting $b \in [0, 100]$, a subject would receive \$12 with probability $(1 - (\mathbf{1}\{\theta = 1\} - b/100)^2)$ and \$6 with complementary probability, where θ equals 1 (-1) when the truth ball is yellow (blue). In the instructions—but not in the main interface—we provided information on the elicitation procedure, phrased as eliciting the probability the truth ball was either yellow or blue, and explained that the procedure was meant to ensure they were incentivized to answer truthfully. To determine payments, we used a report from a single randomly selected period of a randomly selected round. We also asked additional questions on mathematical ability, which were incentivized by providing a \$0.50 reward if they answered correctly a randomly chosen question.

The average compensation was of \$20.02/hour, with subjects spending on average 29 minutes in the experiment. For comparison, this rate is similar to the MTurk experiment of Enke and Graeber (2020), and four times the MTurk average of \$5.00.

Preregistration. Our experiment was registered using the AEA RCT Registry under RCT ID AEARCTR-0003820. The experimental design and recruitment targets were pre-registered, as were our Hypotheses 1, 3a, 4a, 5, and 6. The hypotheses pertaining to response time and belief variance (2 and its variations, 3b and 4b) were introduced subsequently, as feedback we received convinced us they provided evidence for our proposed mechanism.

5. IDENTIFYING RETRACTION FAILURES

We divide our empirical results into two parts. In this section, we demonstrate that retractions fail to correct beliefs and are less effective than new evidence, providing the first identification of retraction failures as a general feature of belief updating. In Section 6, we consider why retractions are less effective and present evidence for our proposed explanation (and against multiple alternatives).

We begin by explaining our empirical strategy in Section 5.1. Section 5.2 briefly compares updating from new observations to that found in existing literature. In Section 5.3, we turn to the main topic of the paper, updating from retractions, testing whether retractions work, and whether people update differently from retractions versus new observations. (Hypotheses 1a and 1b)

5.1. Empirical Strategy

There are two distinct empirical tasks: identifying the effectiveness of retractions for a given history; and aggregating the results across different histories. For both, we lean on the simplicity of our experimental design to make the analysis non-parametric when possible.

5.1.1. Identifying Retraction Effectiveness

To test the effectiveness of retractions and to compare it to new evidence, we perform two distinct comparisons throughout our analysis, corresponding to parts (a) and (b) of Hypothesis 1 and explained visually in Figure 2:

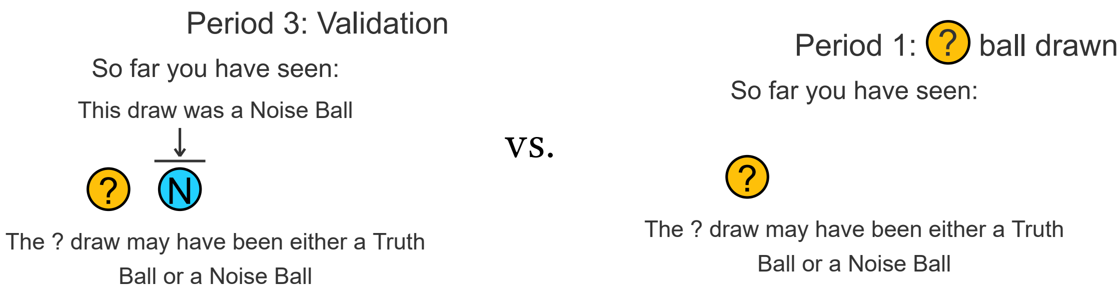
(a) *Testing retraction effectiveness*: Are subjects' beliefs after seeing a retraction the same as if the retracted observation had never been observed in the first place?

$$b(\theta \mid \text{Observations, Retraction of Observation } s_\tau) = b(\theta \mid \text{Observations} \setminus \text{Observation } s_\tau)$$

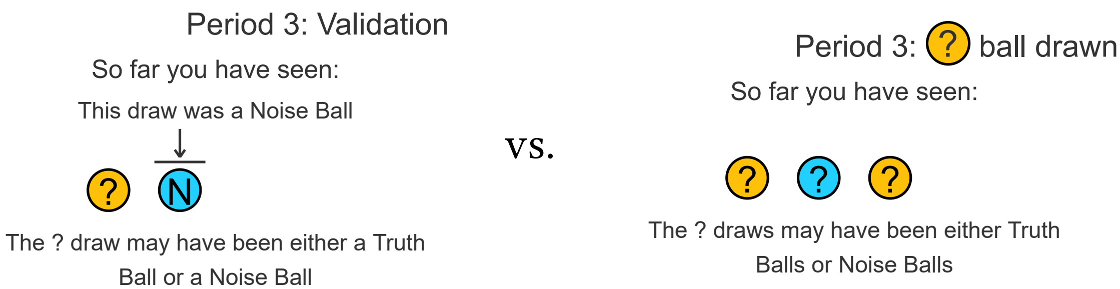
(b) *Comparing retractions to new evidence*: Do subjects update equally from retractions as from equivalent (in terms of Bayesian belief updates) new information?

$$b(\theta \mid \text{Observations, Retraction of Observation } s_\tau) = b(\theta \mid \text{Observations } \cup \text{ New Observation } -s_\tau)$$

To outline our empirical strategy, we introduce some notation. Denote by b the subject's beliefs—the probability they assign to the truth ball being yellow—and by s the signal in question. We treat signals as $+1$ if they favor the belief that the truth ball is yellow (new draws of a yellow



(a) *Do retractions work?* We compare beliefs after a retraction in period t ($\in \{3, 4\}$) to beliefs after an (equivalent) "compressed history" in period $t-2$, the history with the retracted ball removed. The illustrated example compares beliefs after a retraction in period 3 with beliefs in period 1 following a yellow draw.



(b) *Are retractions treated differently from equivalent new observations?* We compare beliefs after a retraction in period t ($\in \{3, 4\}$) to beliefs after an equivalent new observation (of opposite color to the draw which was retracted), also in period t. The illustrated example compares beliefs in period 3 after a retraction of a blue ball and beliefs when the history through period 2 is the same but a yellow ball is draw in period 3.

Figure 2: Illustrative examples to explain the empirical strategy

ball or retractions of a blue ball) and -1 if they favor it being blue. Finally, denote by r a dummy variable indicating whether the signal is a retraction ($r = 1$) or a new observation ($r = 0$).

With this notation in hand, for a specific history, we can both test (a) and (b) with the regression:

$$b = \beta_0 + \beta_1 \cdot r \cdot s. \tag{4}$$

The sample for test (a) comprises beliefs after the retraction as well as when the retracted observation had not been observed to begin with, while for test (b) it comprises beliefs after the retraction and beliefs after a new observation of the opposite sign. The coefficient of interest for both tests is β_1. Under test (a), if β_1 is zero, retractions work: beliefs are as if the retracted signal was never seen; if it is negative, retracted signals continue to influence beliefs. Under test (b), if β_1 is negative, beliefs move less in response to retractions than to equivalent new signals. To give concrete examples, as illustrated in Figure 2, test (a) would compare beliefs in period 3 having observed *(yellow, blue, retraction of the blue)*, to those in period 1 having just observed *(yellow)*;

while test (b) would compare beliefs in period 3 having observed *(yellow, blue, retraction of the blue)*, to those in period 3 having observed *(yellow, blue, yellow)*.

When analyzing retraction effectiveness, test (a), we compare belief reports in levels, while for test (b) we compare effects on beliefs in both levels and changes (first differences), since the test is specifically about how beliefs *change* in response to retractions. We use beliefs as reported by subjects, on a linear scale (0 to 100), except when we analyze biases in belief updating in Section 6.2.1, where we use the log-odds scale to be consistent with existing literature.[14]

5.1.2. Aggregating Results Across Histories, Using Fixed Effects

While we report results disaggregated by case, showing that they are qualitatively consistent across histories, the results are simpler to digest when aggregated. We do so by pooling the sample across histories in suitably modified versions of the above regressions.

The basic identification concern in pooling across histories is using identifying variation which compares updating from retractions in one history to updating from new observations in a different history. We ensure that we are only identifying off within-history variation by using appropriately defined fixed effects. To explain them, denote by H_t the history up to and including period t, that is, the set of all the draws observed as well as the retractions, fixing the order. For the tests of retraction effectiveness, (a), we use fixed effects for what we refer to as a *compressed history*, $C(H_t)$: the history, removing any retracted ball draws as if they had never occurred to begin with, keeping the order fixed. For instance, a history of *(yellow, blue, retraction of the blue)* would be equivalent to *(yellow)*.[15] For the comparisons to new observations, test (b), we include fixed effects at the level of the *sign history*, $S(H_t)$, which is the history without distinguishing whether signals were new observations or retractions. For example, *(blue, yellow, retraction of the blue)* is equivalent to *(blue, yellow, yellow)*. Once we include these fixed effects in the pooled regression, if there have not been retractions in previous periods, then we compare a retraction of the ball of one color to the informationally equivalent new observation of the opposite color, conditional on what happened in all previous periods of the round.

[14] Levels has the advantage that extreme beliefs, near 0 or 100, are not overly inflated; log-odds has the advantage that the experimental signals should lead to a constant change in the log-odds belief, independent of the prior. As we show in Online Appendix B.5.2, our conclusions are robust to relying exclusively on log-odds.

[15] Note that compressed histories do not distinguish between the retracted observation having been drawn in period 1 or period 2. For example, both *(yellow, blue, retraction of the blue)* and *(blue, yellow, retraction of the blue)* have the same compressed history, *(yellow)*.

5.2. Updating from New Observations

As a first step in our analysis, and in part as a test of validity of experimental setting, we examine subjects' belief updating from (non-retracted) new observations using a standard empirical approach in this literature. In the absence of a retraction, the design is similar to many others surveyed by Benjamin (2019) and subjects appear to correctly understand the setting, with reported beliefs tracking Bayesian posteriors closely.[16]

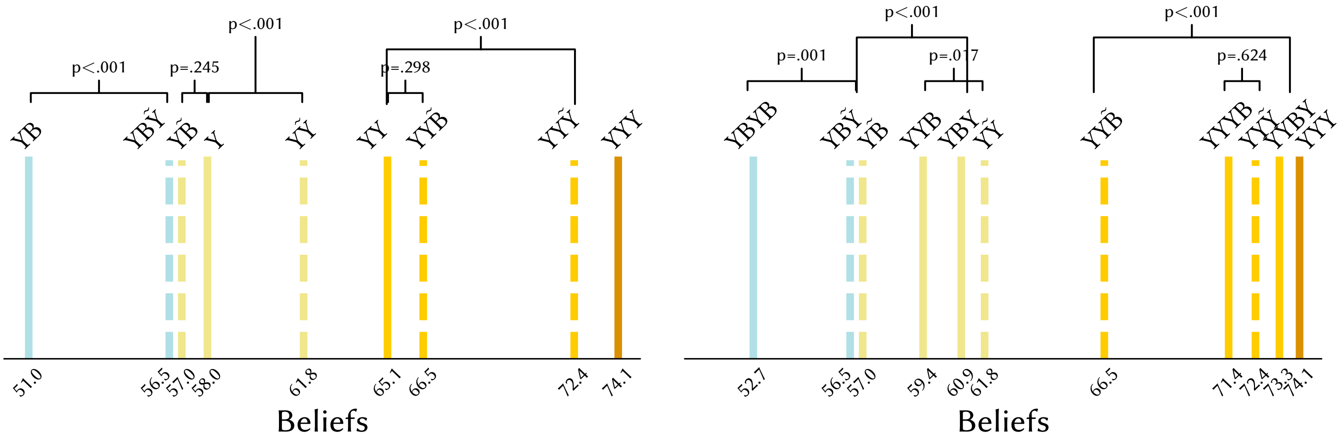
We consider Grether-style (Grether, 1980) regressions—a workhorse model of analysis in this literature—enabling a direct comparison to existing experimental results on belief updating. The results are largely consistent with the main findings from the literature, suggesting that any differences in our subsequent analysis can indeed be attributed to distinct features of retractions. Specifically, we replicate common findings regarding biases in belief updating pertaining to base-rate neglect and confirmation bias. We emphasize that while subjects depart from Bayesian updating, they do so in a way consistent with what one would expect from the literature, and that our theoretical framework nevertheless implies that any additional departure due to retractions cannot be attributed to explanations that are not specific to the nature of the information source. Since this paper focuses on belief updating from retractions, we defer the detailed reporting and discussion of the results to Online Appendix B.1.

5.3. Updating from Retractions (Hypothesis 1)

This section presents our first main findings, on the failure to fully correct beliefs and on the differences in belief updating from retractions as opposed to new observations.

Our first result, and the first central finding of the paper, is the empirical support of Hypothesis 1: retractions are ineffective, in that (a) retracted observations are not fully disregarded (Prior vs. Retraction), and (b) beliefs are less responsive to retractions than to equivalent new signals (Retraction vs. New Draw). Figure 3 depicts mean beliefs across different histories and demonstrates both parts of the hypothesis. In panel (a) we exhibit beliefs following retractions (dashed lines) and beliefs reported in absence of the retract observation (solid lines); panel (b) depicts beliefs following retractions (dashed lines) and contrasts these with beliefs following informationally-equivalent new draws (solid lines). Bars of the same color refer to histories with

[16] In Online Appendix B.3 we show reported beliefs and Bayesian posteriors disaggregated by history (Figure 6) as well as the distance between them (Figure 7).



(a) Prior vs. Retraction

(b) New Draw vs. Retraction

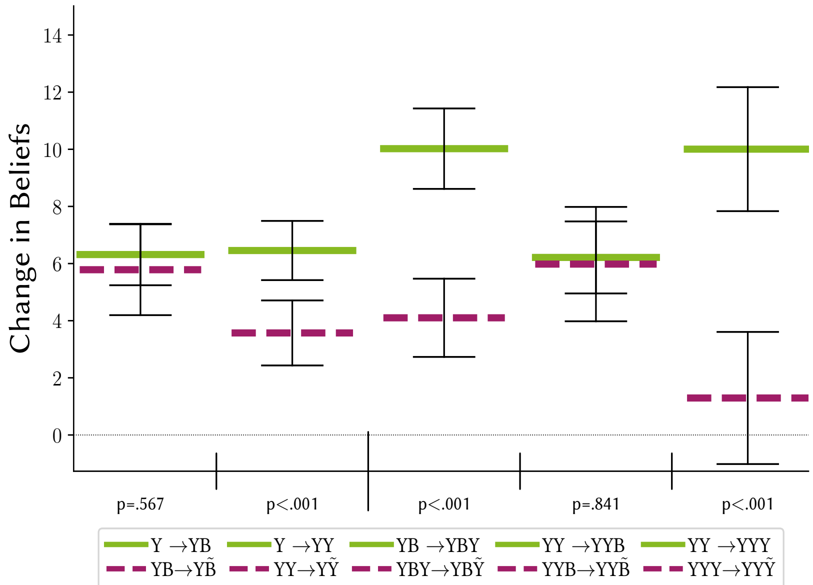
Figure 3: Retractions are Ineffective: Beliefs (Hypothesis 1)

Notes: The figure displays mean reported beliefs, disaggregated by history, where a tilde denotes a retracted observation. Dashed lines indicate histories which end in retractions, solid lines those which do not. Lines of the same color correspond to histories inducing the same Bayesian posterior. Within a color, under Hypothesis 1, mean beliefs will lie to the right after a retracted yellow draw and to the left after a retracted blue draw. p-values for the specific tests are displayed and were obtained by a regression similar to columns (1) and (2) of Table 1 but restricted to the disaggregated histories, using standard errors clustered at the subject level. Belief reports are symmetrized around 50, e.g. $100 - b(B\tilde{Y})$ is treated as $b(Y\tilde{B})$, where a tilde denotes a retracted observation. The sample paths do not condition on sequence order: e.g. $Y\tilde{B}$ and $\tilde{B}Y$ are bundled together. The sample consists of subjects in the baseline treatment (beliefs elicited each period).

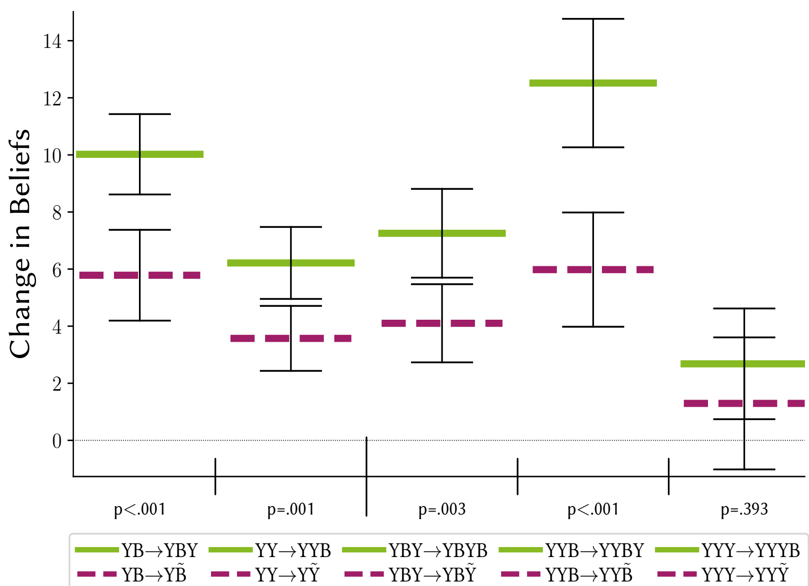
identical Bayesian posteriors; p-values refer to the statistical significance of regression coefficients from regressions following tests outlined in Section 5.1 but restricting the sample to observations with identical Bayesian posteriors. We find overwhelming evidence for retraction ineffectiveness across the different histories.

Figure 4 confirms this when looking at *changes* in beliefs: the change in beliefs following a retraction is smaller than the changes in beliefs both when (a) the subsequently-retracted observation was originally observed, and (b) when, instead of a retraction, subjects observe an equivalent new observation. Again, in panel (a) we compare inference—change in beliefs, in percentage points—from retractions (dashed lines) to inference from the observation which was retracted (solid lines) by conditioning on histories with equivalent Bayesian posteriors; panel (b) does the same but with retractions (dashed lines) and new observations (solid lines). Results point to significant underinference from retractions *relative to* observations, that is, direct information about the state.

We pool these results across different histories in Table 1. Column (1) is a test of (a) and



(a) Previous Draw vs. Retraction



(b) New Draw vs. Retraction

Figure 4: Retractions are Ineffective: Changes in Beliefs (Hypothesis 1)

Notes: The figure compares the change in beliefs following the retraction of an observation to (a) the change in beliefs when the observation was first drawn, and (b) the change in beliefs following an equivalent new observation. Belief reports are symmetrized around 50, e.g. $-(b(B\tilde{Y}) - b(BY))$ is treated as $(b(Y\tilde{B}) - b(YB))$; equivalently, we normalize the direction in which the updating should occur by considering $\Delta b_t \cdot s_t$. Values are changes in reported beliefs in percentage points. The sample paths do not condition on sequence order: e.g. $Y\tilde{B}$ and $\tilde{B}Y$ are bundled together. The sample consists of subjects in the baseline treatment (beliefs elicited each period). The whiskers denote 95% confidence intervals using standard errors clustered at the subject level; p-values were obtained by auxiliary regressions similar to column (3) of Table 1, but restricting to the disaggregated histories.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	0.201	-0.167	-0.351	-0.242
	(0.275)	(0.368)	(0.355)	(0.363)
Retracted Signal ($r_t \cdot s_t$)	-3.134***	-3.628***	-3.701***	-3.316***
	(0.601)	(0.726)	(0.670)	(0.675)
Signal (s_t)	–	–	–	8.658***
				(0.510)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.34	0.34	0.18	0.15
Observations	22578	22578	22578	9074

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Retractions are Ineffective (Hypothesis 1)

Notes: This table tests compares whether retractions return beliefs to what they would have been had the retracted signal never been observed, and compares their effectiveness relative to new direct information. The sample includes beliefs of subjects in the baseline treatment (beliefs elicited each period). Column (1) tests part (a) of the hypothesis, whether retractions work, by comparing beliefs after a retraction to beliefs after the equivalent compressed history. The outcome is the beliefs in period t, $b_t \in [0, 100]$. In the case of a retraction, s_t is the opposite sign of the original observation being retracted in round t (+1 if an earlier -1 signal is retracted, -1 if an earlier +1 signal is retracted). The regression includes fixed effects for the compressed history of draws. Columns (2) to (4) test part (b) of the hypothesis, whether people update less from retractions compared to equivalent new observations. The specifications include fixed effects for the sign history. In column (2), the outcome is the beliefs in period t, b_t. In columns (3) and (4), the outcome is the first difference in beliefs. Column (4) uses lagged sign history fixed effects to enable us to compare the magnitude of $r_t \cdot s_t$ to s_t, which is otherwise absorbed by the fixed effects. In columns (1)-(3), the sample excludes cases in which the truth ball is disclosed and in which there was a retraction in the past; column (4) further restricts to periods 3 and 4.

corresponds to the following regression:[17]

$$b_t = \beta_1 \cdot r_t \cdot s_t + \beta_2 \cdot r_t + F_{C(H_t)}, \tag{5a}$$

where we exclude from the sample cases in which the truth ball is disclosed or in which there was a previous retraction. As explained in Section 5.1.2, controlling for compressed history fixed

[17]The coefficient on r, which is added when we aggregate across histories, has no meaning in itself: it identifies whether beliefs are on average shifted toward yellow when retractions occur, given the fixed effects. The coefficient will depend not only on how the realized frequency of blue and yellow observations compares to that of retractions, but also on which one is the base group for the fixed effects.

effects $F_{C(H_t)}$ compares, for example, the beliefs after observing $(s_1, s_2, n_2 = 1)$ to those reported when only observation s_1 was seen.

Columns (2)-(4) test (b) and correspond to variants of the following regression:

$$b_t = \beta_1 \cdot r_t \cdot s_t + \beta_2 \cdot r_t + \beta_3 \cdot s_t + F_{S(H_t)}, \tag{5b}$$

where we again exclude from the sample cases in which the truth ball is disclosed or in which there was a previous retraction. Controlling for sign history fixed effects, $F_{S(H_t)}$, means we compare for example beliefs reported after $(s_1, s_2, n_2 = 1)$ to those reported after $(s_1, s_2, s_3 = -s_2)$. In column (2) the dependent variable is belief levels, whereas in columns (3) and (4) it is *change* in beliefs. Column (4) uses less stringent fixed effects—those for *lagged* signed history $F_{S(H_{t-1})}$—so that the signal term s_t is not absorbed by the fixed effects, but only includes periods 3 and 4. This enables us to benchmark the differential effect of retractions, β_1, by comparing it with the effect of new observations, β_3.

The key finding for both tests is that the differential effect of retractions on beliefs, β_1 the coefficient on $r_t \cdot s_t$, is negative and consistent in magnitude across all of the specifications we study. Retractions are treated *differently*, and in particular as if they were less informative than equivalent new observations. To quantify this effect, a simple comparison shows that beliefs move approximately one-third less when information is in the form of a retraction. This can be seen from column (4) of Table 1, by comparing the coefficient on the retracted signal—the interaction term between the signal and the retraction variables—to the coefficient on the signal variable itself. Performing this back-of-the-envelope calculation in other ways, for example by dividing the coefficient on $r_t \cdot s_t$ in column (3) by the average update from a new observation in the corresponding sample, consistently finds that beliefs update around $1/3$ from retractions relative to new observations.

6. WHY DO RETRACTIONS FAIL?

We now turn to why retractions are not effective, and why they affect beliefs less than equivalent new observations. If the ineffectiveness of retractions was due to biases in updating from new observations, by Proposition 1, there would be no difference in updating from new observations versus from retractions. Having demonstrated that, in fact, retractions are treated as less informative than an otherwise equivalent new observation (Hypothesis 1b), our starting point is that any explanation should be germane to retractions themselves—established biases such as confirmation bias cannot explain our results. We divide our analysis of possible mechanisms into two parts. First, we provide evidence that retractions, being are information about information, are fundamentally more complex and harder to process. Second, we discuss alternative mechanisms which could plausibly generate our results, and show that they do not.

6.1. Retractions are Harder to Process

We propose and find support for a mechanism which could explain why retractions are less effective: retractions are simply harder to process. Our design made retractions as simple as possible, rendering retracting a signal equivalent to the intuitive benchmark of deleting it, as well as to an opposite observation. Yet, despite the informational equivalence between observations and retractions, the former are simply information, while retractions are information *about* information, indicating how to interpret past observations. This qualitative difference implies that retractions necessitate conditional reasoning, which may make it harder to assess their informational content.

We explore hypotheses suggested by the theoretical model in Section 2, which posit that the diminished effectiveness of retractions is caused by a corresponding increase in the cognitive difficulty of updating (measured by σ_ζ^2). We rely on two commonly used proxies for cognitive imprecision: decision times and variability in responses—we refer to Section 2 for a discussion. First, we test whether retractions induce longer decision times and greater belief variance (Hypothesis 2). Second, we posit and argue that retractions of less recent observations are relatively harder to process (Hypothesis 3) and that inferring from new observations *following* retractions is also harder (Hypothesis 4). We test and verify that our behavioral markers in both cases are aligned with our conjecture—i.e. decision times are longer and belief variance is higher for situations conjectured to be more complex—and show that indeed subjects infer less in the presumed more complex situation.

6.1.1. Retractions, Longer Decision Times and Greater Belief Variance (Hypothesis 2)

The idea that retractions are more cognitively taxing motivates our hypothesis that they result in greater cognitive imprecision and thus induce longer decision times and greater variability in responses, as discussed above.

To test whether decision times are longer when subjects are faced with a retraction (Hypothesis 2a), we use an identification strategy similar to the one used to test the effects of retractions on

	Prior vs. Retraction		New Draw vs. Retraction	
	(1)	(2)	(3)	(4)
	log(dt$_t$)	V(b$_t$ $\mid h_t$)	log(dt$_t$)	V(b$_t$ $\mid h_t$)
Retraction (r_t)	0.053***	128.1***	0.101***	71.7***
	(0.016)	(20.300)	(0.015)	(22.119)
Mean Decision Time (secs)	6.674	–	6.674	–
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R-Squared	0.01	0.02	0.02	0.02
Observations	22578	3030	22578	3030

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Retractions are Harder (Hypothesis 2)

Notes: This table tests if retractions induce longer decision times and greater belief variance. Columns (1) and (3) test if retractions induce longer decision times, measured in log-seconds (Hypothesis 2a). The sample includes decision times of subjects in the baseline treatment (beliefs elicited each period), excluding cases in which the truth ball is revealed and those in which there was a retraction in the past. Column (1) compares decision time after a retraction to decision time after the equivalent compressed history. Column (3) compares decision time after a retraction to decision time after an equivalent new draw. Columns (2) and (4) test whether retractions induce higher belief volatility (Hypothesis 2b). Column (2) compares belief variance following a retraction to belief variance at equivalent histories in which the retracted signal was not drawn; the dependent variable is the sample variance of beliefs of a given subject, conditional on permuted compressed histories and on whether a retraction occurred. Column (4) compares belief variance following a retraction to belief variance following an equivalent new draw; the dependent variable is the sample variance of beliefs of a given subject, conditional on permuted sign histories and on whether a retraction occurred.

belief updating (Section 5.1). Specifically, in Table 2, we estimate two versions of

$$\log(dt_t) = \beta_1 \cdot r_t + F \tag{6a}$$

where dt_t is the subjects' decision time in seconds. F corresponds to either compressed history fixed effects or sign history fixed effects, as in columns (1) and (2)-(4) of Table 1 and under the same sample restrictions. In the first case, we compare decision times following a retraction with those at histories in which the retracted observation was never drawn. In the second, we compare decision times following a retraction versus an informationally equivalent new observation.

We pursue an analogous strategy to identify the effect of retractions on belief variance and test Hypothesis 2b. To test if retractions increase belief variance relative to histories in which the retracted observation was never drawn, we calculate the sample belief variance at the subject-level conditional on the compressed history and on whether a retraction was observed, $\text{Var}(b_t \mid C(H_t), r_t)$. We treat compressed histories that are the same up to permutations as the same compressed history so as be able to estimate within-subject belief variance at a given (permuted) compressed history. To we test whether belief variance following a retraction is greater than at informationally equivalent histories in which the retracted observation was never drawn, we estimate the following equation:

$$\text{Var}(b_t \mid C(H_t), r_t) = \beta_1 \cdot r_t + F_{C(H_t)}. \tag{6b}$$

Analogously, to compare belief variance following a retraction versus an equivalent new observation, we calculate the sample belief variance, for each subject, conditional on the sign history $S(H_t)$ (allowing for permutations) and on the occurrence of a retraction in period t, and estimate

$$\text{Var}(b_t \mid S(H_t), r_t) = \beta_1 \cdot r_t + F_{S(H_t)}. \tag{6c}$$

We note that, by conditioning on permutations of compressed/sign histories, we hold fixed Bayesian posteriors; i.e. permutations of a compressed/sign history are informationally equivalent. For notational simplicity, we write $V(b_t \mid h_t)$ to denote the subject-level sample variance conditional on (permuted) compressed/sign histories as described above.

The results, in Table 2, confirm both Hypotheses 2a and 2b. Specifically, column (1) shows that subjects take approximately 5% longer reporting beliefs following a retraction when compared to cases in which the retracted observation was never drawn. Comparing retractions with equivalent new draws, belief reporting following a retraction takes 10% longer (column (3)). Columns (2) and (4) provide an analogous comparison for belief variance estimated at the subject level, which retractions increase by over one third. In both cases, we see that belief variance increases following a retraction. In Appendix B.5.8, we provide an alternative test for whether variance of beliefs is higher for retractions than for new observations, whereby we disaggregate by (permuted) compressed and sign histories; the results are consistent. In Appendix B.5.7, we show results on response times remain valid when controlling for the round number and its interaction with retractions. While our results show subjects take less time in later rounds, the increase in decisions time caused by retractions remains consistent in later rounds, when subjects have had more experience observing retractions.[18]

[18]Note that subjects are fully informed they may see a retraction prior to any round where they do, and the interface is as similar as possible for new draws and retractions; hence we do not find it surprising we do not detect a difference depending on whether subjects has seen more retractions in the past.

Our results suggest that retractions are not only treated differently, they are also harder to process. In line with the literature on cognitive imprecision, one interpretation consistent with our results is that such increased complexity is reflected in a noisier perception of the informativeness of a retraction relative to direct information about the state of the world.

6.1.2. Harder Retractions are Harder (Hypothesis 3)

We exploit the fact that our design entails dynamic arrival of information to distinguish retractions according to whether or not they refer to the last observation. Specifically, we argue that retractions of observations received earlier may induce a more complex reevaluation of previously observed signals, relative to retractions of observations received more recently. When, at time t, the observation received at $t-1$ is retracted, then subjects need but to revert to the belief they held at $t-2$, that is, before receiving that observation. Hence a natural conjecture is that retractions are more effective when they to refer to information that was just received (Hypothesis 3a), in which case subjects only need report their beliefs from the previous period. Moreover, if retractions of more recent observations are then less complex and insofar as decision time and belief variance proxy for cognitive complexity, lower decision times and belief variance in cases in which retracted observations occurred in the previous period (Hypothesis 3b).

In columns (1) and (4) of Table 3, we report the same basic specifications described in Section 5.3, tests of (a) retractions and (b) retractions versus new information, but with the addition of an indicator variable for whether the last signal observed was retracted (rl_t), as well as its interaction with the signal itself. We find greater effectiveness of retractions when these correspond to the most recently received ball draw: it is easier to disregard a piece of information if it arrived more recently. However, subjects still fail to fully disregard retracted signals, even when they are of the most recent draw, as reflected by the sum of the coefficients on $r_t \cdot s_t$ and $rl_t \cdot s_t$ being negative (p-value$= .021$) and over 1/3 of the size of $r_t \cdot s_t$.

Analogously, we expand the analysis of decision times and of (conditional sample) variance of beliefs from Table 2 to also account for whether the retracted observation corresponds to the last draw or not. Columns (2) and (5) add the indicator for whether the last signal observed was retracted (rl_t) to equation (6a). Columns (3) and (6) mimic the specifications in equation (6b) and (6c) by conditioning not only on permuted compressed/sign history and the occurrence of a retraction, but also on whether the retraction refers to the last draw. In line with greater

	Prior vs. Retraction			New Draw vs. Retraction		
	(1)	(2)	(3)	(4)	(5)	(6)
	b_t	$\log(dt_t)$	$V(b_t \mid h_t)$	b_t	$\log(dt_t)$	$V(b_t \mid h_t)$
Retraction (r_t)	0.630	0.088***	131.1***	-0.158	0.119***	77.5***
	(0.412)	(0.018)	(24.558)	(0.438)	(0.018)	(26.426)
Retracted Signal ($r_t \cdot s_t$)	-4.341***	–	–	-4.258***	–	–
	(0.685)			(0.800)		
Last Draw Retracted (rl_t)	-0.981	-0.080***	-80.828**	-0.029	-0.042*	-76.454**
	(0.662)	(0.019)	(33.165)	(0.697)	(0.021)	(33.968)
Retracted Signal	2.737***	–	–	1.436**	–	–
x Last Draw Retracted ($rl_t \cdot s_t$)	(0.678)			(0.641)		
Mean Decision Time (secs)	–	6.674	–	–	6.674	–
Compressed History FEs	Yes	Yes	Yes	No	No	No
Sign History FEs	No	No	No	Yes	Yes	Yes
R-Squared	0.34	0.01	0.02	0.34	0.02	0.02
Observations	22578	22578	3295	22578	22578	3295

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Harder Retractions are Harder (Hypothesis 3)

Notes: This table tests if retractions of less recent observations: are less effective (columns (1) and (4)), result in shorter decision times (columns (2) and (5)), and induce lower belief variance (columns (3) and (6)). The sample is restricted to subjects in the baseline treatment (beliefs elicited each period). The dependent variable in columns (1) and (4) is the beliefs in period t, $b_t \in [0, 100]$. In the case of a retraction, s_t is the opposite sign of the original observation being retracted in round t (+1 if an earlier -1 signal is retracted, -1 if an earlier +1 signal is retracted). In columns (2) and (5), the dependent variable corresponds to the decision time at a given period. The dependent variable in columns (3) and (6) is the sample variance of beliefs of a given subject, conditional on on whether or not a retraction occurred, on whether the last observation was retraction, and on the permuted compressed history (column (3)) or the permuted sign history (column (6)). Columns (1) and (4) exclude cases in which the truth ball is disclosed and in which there was a retraction in the past; columns (2) and (5) consider decision times under the same conditions.

effectiveness, we further observe that belief reporting is starkly faster when the retraction refers not to an earlier but to the last draw—columns (2) and (5) of Table 3—and also that retractions of more recent observations induce lower belief variances—columns (3) and (6) of Table 3.

To summarize, our results suggest that more recent observations are easier to retract than more distant observations: their retractions are cognitively less demanding and also more effective—albeit less than obtaining informationally equivalent direct information.

6.1.3. Updating After Retractions (Hypothesis 4)

To conclude this section, we consider the effect that retractions have on updating from subsequent new evidence. To our knowledge, this is the first time that data of this kind is collected and analyzed, and, therefore, existing literature provides little guidance. Our design precludes any consideration of drawing inferences on the credibility of the source following a retraction. Our posited mechanism, however, suggests that if a retraction is more difficult to process, it may be more difficult to update following a retraction.

We first test if beliefs change less in response to a signal following a retraction by estimating

$$\Delta b_t = \beta_1 \cdot r_{t-1} \cdot s_t + \beta_2 \cdot r_{t-1} + F_{S(H_t)} \tag{7a}$$

where r_{t-1} is an indicator for whether a retraction occurred in the previous period. We consider all beliefs from all periods, except cases when the truth ball is disclosed or when there are multiple retractions within a round. The use of sign history fixed-effects $F_{S(H_t)}$ allows us to identify the effect of a retraction in the previous period on belief updating. Our dependent variable of interest is the *change* in beliefs Δb_t and not level beliefs b_t since observing an effect could be due to the effect of retractions on updating at $t-1$ and not on subsequent updating.

To gauge if updating from new observations after retractions is harder, we use an analogous identification strategy and estimate

$$\log(dt_t) = \beta_1 \cdot r_{t-1} + F_{S(H_t)}. \tag{7b}$$

Finally, we rely on the approach as described in Section 6.1.1 to test if previous retractions affect belief volatility. In particular, we calculate for each subject the sample variance of their beliefs conditional on whether or not a retraction occurred in the previous period and on the (permuted) sign history; we denote such quantity by $V(b_t \mid h_t)$. Then, using $V(b_t \mid h_t)$ as a dependent variable, we estimate a version of equation (7b), where fixed effects refer to permuted sign history.

The results are presented in Table 4. Column (1) indicate that subjects do in fact infer less from new observations following a retraction, with the coefficient on $r_{t-1} \cdot s_t$ being significantly negative. Moreover, having had a retraction in the previous period does significantly increase decision time by 8% (column (2)), indicating greater cognitive imprecision in evaluating available evidence. Column (3) shows that belief variance increases, further corroborating this assessment. To conclude, retractions make subsequent updating from news observations less responsive and more cognitively demanding, consistent with retractions themselves being harder to process.

	New Draw vs. Retraction		
	(1)	(2)	(3)
	Δb$_t$	log(dt$_t$)	V(b$_t$ $\mid$ h_t)
Retraction (r_{t-1})	0.794	0.083***	136.0***
	(0.805)	(0.025)	(40.526)
Retracted x Signal ($r_{t-1} \cdot s_t$)	-1.881**	–	–
	(0.782)		
Mean Decision Time (secs)	–	6.675	–
Compressed History FEs	No	No	No
Sign History FEs	Yes	Yes	Yes
R-Squared	0.18	0.02	0.03
Observations	21270	21270	2611

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Updating After Retractions (Hypothesis 4)

Notes: This table tests if processing new observations after retractions results is more difficult, resulting in subjects updating less from new observations (column (1)), in longer decision times (column (2)), and greater belief variance (column (3)). The sample includes beliefs of subjects in the baseline treatment (beliefs elicited each period). In columns (1)-(2), the sample excludes cases in which the truth ball is disclosed and in which there was a retraction in period 4. Column (3) compares belief variance after a retraction to belief variance after an equivalent new draw; the dependent variable is a subject's sample variance of beliefs, conditional on permuted sign histories and on whether a retraction occurred in the previous period.

6.2. Alternative Explanations

In this section, we consider alternative—but not necessarily competing—explanations for why retractions are ineffective in correcting beliefs and less effective than new signals. First, we examine if biases in updating from retractions are nevertheless similar to those in updating from new observations, and whether such similarities could underlie retraction failures. Second, we test whether retraction ineffectiveness is driven by retracted evidence having been actively used by the subjects. We then conclude with a summary of the mechanisms which we are able to rule out.

6.2.1. Similar Updating Biases (Hypothesis 5)

According to Proposition 1, any explanations that do not treat retractions differently—such as confirmation bias—cannot explain why retractions fail. Here we seek to determine whether, despite their being treated differently, updating from retractions is nevertheless affected by biases akin to those affecting updating from new observations, and whether similarity in these biases

	All Periods		Period 3	
	(1)	(2)	(3)	(4)
	l_t	l_t	l_t	l_t
Prior (l_{t-1})	0.834***	0.800***	0.904***	0.839***
	(0.037)	(0.037)	(0.051)	(0.051)
Signal (s_t)	1.126***	0.998***	1.647***	1.314***
	(0.071)	(0.072)	(0.135)	(0.142)
Signal Confirms Prior ($s_t \cdot c_t$)	–	0.417***	–	0.705***
		(0.135)		(0.246)
Retraction (r_t)	-0.033	-0.030	-0.034	-0.026
	(0.021)	(0.021)	(0.033)	(0.034)
Retraction x Prior ($r_t \cdot l_{t-1}$)	0.019	0.070*	-0.093*	-0.002
	(0.040)	(0.042)	(0.053)	(0.051)
Retracted Signal ($r_t \cdot s_t$)	-0.768***	-0.541***	-1.286***	-0.825***
	(0.092)	(0.097)	(0.156)	(0.160)
Retraction x Signal Confirms Prior ($r_t \cdot s_t \cdot c_t$)	–	-0.675***	–	-1.051***
		(0.173)		(0.241)
R-Squared	0.44	0.44	0.43	0.43
Observations	22578	22578	6081	6081

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Similar Updating Biases (Hypothesis 5)

Notes: This table tests whether biases in belief updating are different when updating from retractions. It estimates the standard specifications in Table 8, but interacting its terms with whether the signal was a retraction. The sample consists of all elicited beliefs of subjects in the baseline treatment (beliefs elicited each period), excluding those elicited after a verification, within a given round. Columns (1) and (2) consider all periods; columns (3) and (4) restrict the sample to period 3. The outcome is the log-odds of beliefs in period t, l_t. s_t is the signal in round t (+1 or -1, multiplied a constant factor of Bayesian updating such that the coefficient on s_t would be 1 under Bayesian updating). $c_t := \mathbf{1}\{\text{sign}(l_{t-1}) = \text{sign}(s_t)\}$ denotes whether the signal at t confirms the prior at $t-1$.

could explain retraction failures.

Specifically, we utilize the conventional Grether-style (Grether, 1980) specifications in the literature analysing other deviations from Bayesian updating in similar experiments. These rely on log-odds specifications—discussed in Section 2—that take the following form:

$$l_t = \beta_0 + \beta_1 \cdot l_{t-1} + \beta_2 \cdot s_t \cdot K \tag{8a}$$

and $$l_t = \beta_0 + \beta_1 \cdot l_{t-1} + \beta_2 \cdot s_t \cdot K + \beta_3 \cdot s_t \cdot K \cdot c_t \tag{8b}$$

where t is the period, l_t is the log-odds of the beliefs reported at t,[19] s_t is the signal in round t (+1 or -1), $c_t := \mathbf{1}\{\text{sign}(l_{t-1}) = \text{sign}(s_t)\}$ is an indicator function that equals 1 when the signal at t confirms the prior at $t-1$, and $K > 0$ is a constant factor of Bayesian updating (the log-likelihood of the signal).

In the above specifications, a Bayesian decisionmaker would exhibit a coefficient $\beta_2 = 1$, and $\beta_2 < 1$ or > 1 would imply, respectively, under- and over-inference from signals. Similarly, Bayesian updating would entail $\beta_1 = 1$, while a common finding is that belief updating exhibits base-rate neglect, with $\beta_1 < 1$. Finally, confirmation bias is captured by $\beta_3 > 0$, while instead a Bayesian would have $\beta_3 = 0$. As discussed in Sections 5.2, we estimate these specifications restricting the sample to histories in which there is no retraction and we finding patterns standard in the literature—results are reported in Section B.1.

We here return to these regression specifications and fully interact them with the retraction variable, r_t, corresponding to whether the signal was in the form of a retraction:

$$l_t = \beta_0 + \beta_1 \cdot l_{t-1} + \beta_2 \cdot s_t \cdot K + \beta_3 \cdot s_t \cdot K \cdot c_t + \\ + r_t \cdot [\gamma_0 + \gamma_1 \cdot l_{t-1} + \gamma_2 \cdot s_t \cdot K + \gamma_3 \cdot s_t \cdot K \cdot c_t] \tag{9}$$

The inclusion of the interactions allows us to detect how previously documented deviations from Bayesian updating vary, depending on whether or not the signal is a retraction. In other words, they provide a flexible functional form in order to capture the effect of retractions as discussed in Section 2.

The results can be found in Table 5. A striking pattern emerges: when updating from new draws subjects (slightly) overinfer from signals ($\beta_2 \geq 1$) and do more so when signals confirm the prior ($\beta_3 > 0$); in contrast, when updating from retractions they *under*-infer ($0 < \beta_2 + \gamma_2 < 1$) and exhibit *anti*-confirmation bias ($\beta_3 + \gamma_3 < 0$).[20] In sum, belief updating from retractions exhibits the opposite biases to updating from new draws, a conclusion which is robust across all specifications. This strengthens the conclusion that retractions are treated differently from new signals, inasmuch as the behavioral responses to retractions are not simply accentuating

[19] All tables involving log-odds of beliefs treat $b_t = 100$ and $b_t = 0$ respectively as $b_t = 100 - \delta$ and $b_t = \delta$. We chose $\delta = 0.1$ so as to avoid biasing the regression with extreme outliers. The results are robust to varying δ and to dropping subjects that answer $b_t \in \{0, 100\}$.

[20] We conduct F-tests to analyze the statistical significance of such observations: β_2 is not significantly different from 1 in column (2) (p-values$=.974$), and significantly larger than 1 in the remaining columns; $\beta_2 + \gamma_2$ is always significantly smaller than 1 and larger than 0 (p-value$< .001$ in all cases); $\beta_3 + \gamma_3$ is always negative and significantly different from zero (p-value$< .001$ for column (3) and .011 for column (4)).

pre-existing biases; in fact, retractions induce opposite biases in belief reporting behavior.

6.2.2. Retracting Used Evidence (Hypothesis 6)

Here we consider another potential explanation for why retractions fail: that it is difficult to disregard evidence that has been actively used, as might be suggested by explanations based on cognitive dissonance. We test this hypothesis by comparing updating from retractions when beliefs have already been elicited to when they have not, by contrasting belief reports across our between-subject treatments: the baseline treatment—in which beliefs are elicited every period within a round—and the single-elicitation treatment—in which beliefs are elicited only at the end of the round.

The results from this comparison are documented in Table 6. The specifications correspond to equations (5a) and (5b) which we described in Section 5.3, with the addition of the single-elicitation treatment as an interaction term.[21] The result is a null result: having acted upon a piece of information does not change the effect of it being retracted on belief updating. As before, beliefs move in the directions of signals, but less so for retractions relative to new signals. While this does not imply that retractions are as (in)effective when individuals acted upon past information in other contexts, it does strengthen our conviction that our results are not due to design details.

6.2.3. Explanations for Retraction Failures Ruled out by our Design

We conclude by taking stock of alternative explanations for retraction failure that we rule out based on the design itself.

First, we recall from the introduction that our use of an balls-and-urns design was motivated by our desire to tie retraction failures documented in Section 5.3 to belief updating itself, minimizing the role of explanations related to particular domains (e.g., scientific understanding or political preferences). In particular, the fact that motivated reasoning is often at play in political domains might suggest it plays an important role in the limited effectiveness of retractions; while this could magnify the effect we find, we find this effect even without motivation. We emphasize that the paradigm we use has the advantage of allowing us to quantify objectively correct beliefs, which is difficult or impossible to do in domains where beliefs are subjective or, perhaps more

[21] When beliefs are elicited only at the end of each round, it is not possible to obtain the first difference in beliefs and there is therefore no reasonable way to estimate columns (3) and (4) of Table 1.

	Prior vs. Retraction	New Draw vs. Retraction
	(1)	(2)
	b_t	b_t
Final (Fin_t)	0.175	0.143
	(0.938)	(0.997)
Retraction (r_t)	-0.048	-0.230
	(0.326)	(0.400)
Retracted Signal ($r_t \cdot s_t$)	-2.404***	-3.658***
	(0.622)	(0.712)
Final x Retraction ($Fin_t \cdot r_t$)	–	0.049
		(0.754)
Final x Retracted Signal ($Fin_t \cdot r_t \cdot s_t$)	0.132	0.154
	(0.999)	(0.995)
Compressed History FEs	Yes	No
Sign History FEs	No	Yes
R-Squared	0.21	0.31
Observations	11213	9920

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Retracting Used Evidence (Hypothesis 6)

Notes: This table tests whether updating from retractions is different if beliefs have previously been elicited before a signal is retracted. It estimates specifications in Table 1, but interacting its terms with an indicator for whether the subject was in the single-elicitation treatment (Fin_t). The sample includes all subjects, both those in the baseline treatment (beliefs are elicited each period) as well as those in the final elicitation treatment (beliefs are elicited only at the end of the round), but restricted to periods 1 to 3 and to cases in which the truth ball is not disclosed. Column (1) restricts to period 1 and to period 3 when there is a retraction, interacting the specification from column (1) in Table 1—are retractions effective—with a dummy for being in the single-elicitation group ($Fin_t \cdot r_t$ and $Fin_t \cdot s_t$ are spanned by the other controls and hence omitted, since period 3 is only in the sample when it is a retraction, making Fin_t = $Fin_t \cdot r_t$ within the sample). Column (2) restricts to period 3, interacting the specification from column (2) in Table 1—is updating from retractions different from updating from new observations—with a dummy for being in the single-elicitation group.

problematically, not concretely defined. Issues of whether retractions lead to questioning the source's reliability, while interesting in their own right, are also precluded in our setting.

Second, as Proposition 1 demonstrates, only explanations specific to retractions can rationalize retraction failures. Indeed, we design our experiment such that we can compare retractions to other pieces of equivalent information. We can thus distinguish retraction failures from any explanation that applies to all forms of information processing and belief updating, such as confirmation bias.

Third, since our design nests the classic balls-and-urns setup, it allows us to replicate and compare our findings with the existing literature on biases in belief updating, and show that the failure of retractions is a distinct phenomenon. Proposition 1 shows that the biases in updating from new information cannot by themselves explain the failure of retractions, as otherwise there would be no difference in updating from retractions and equivalent new observations. Our results studying such biases further show that they are also *qualitatively* different for retractions as compared to new observations, as shown above in Section 6.2.1: biases in updating from retractions are not simply accentuated versions of pre-existing biases.

Fourth, we tested whether retractions fail due to subjects having used the retracted evidence in the past. As reported above in Section 6.2.2, we fail to detect any effect difference on the (in)effectiveness of retractions when compared to a case in which retracted evidence was never acted upon prior to its retraction. Finally, we note that, as the subjects have access, at all times, to the whole history of observations and (if any) retractions, we rule out memory issues related to imperfect recall of previous draws of which observation is retracted.

In sum, while Section 6.1 tests and provides support for one mechanism—that retractions are treated as more complex than new direct evidence—here we have ruled out a plethora of other mechanisms either by testing them explicitly or by design.

7. ROBUSTNESS CHECKS

We subjected our results to a battery of robustness checks. In this section, we address two possible concerns: that results are driven by subjects failing to understand the setting, and that they are driven by a small fraction of subjects instead of being a general feature in our sample.

7.1. Subject Screening

We strove to ensure that our results were not driven by inattentive subjects. While behavior of participants in our choice of subject pool (Amazon Mechanical Turk) has been shown to approximate well a representative population sample, it is also the case that behavior is 'noisy' relative to a traditional laboratory subject pool (Snowberg and Yariv, 2021; Gupta et al., 2021).

We went to lengths to filter out bots and overly inattentive subjects at the start of the experiment. Specifically, we took four main steps in order to ensure that our subject pool was of high quality.

First, we included captchas throughout the experiment in order to filter out bots. Second, we included comprehension questions in the instructions which subjects needed to answer correctly in order to proceed with the experiment. The questions summarized the key points the subjects needed to understand, and would have been very difficult to answer correctly without having understood the instructions. While unincentivized, the majority of the subjects answered all questions correctly on the first try (55%), and 90% answered correctly by the second try—with uniform random guesses, the probability of answering all correctly on first try would be lower than 1%. Third, as detailed in Section 4.3, we used a payment scheme which involved a high baseline and reward pay. Fourth, we restricted our study to be held only during business hours (Eastern Standard Time), and we restricted eligibility to US adults and precluded the possibility of repeating the experiment.

These quality checks were important for us to be able to meaningfully test our hypotheses. Excessively noisy answers would have attenuated our results: while a subject answering 50-50 to everything would not be a Bayesian, they would also demonstrate no differential updating from retractions. It was also important that subjects understood retraction should *not* be treated as evidence for the opposite state. Misinterpreting the instructions in this way would suggest retractions should be treated as *more* informative than new information, again working against us finding evidence for our hypothesis.

7.2. Subject Understanding

We also check robustness to excluding subjects based on different measures of inattentiveness. The results are robust, and in fact slightly stronger, when restricting the sample to those subjects who appear attentive, as defined in three different ways. First, using the (unincentivized) comprehension questionnaire that followed the presentation of the instructions, we restrict our sample to subjects who answered all questions correctly on their first try, who account for a majority. Second, we further restrict the sample to subjects who, when the state is revealed, correctly report that they know the state. Third, we remove subjects whose belief reports are excessively noisy, which we define as updating in the opposite direction to the signal more than 10% of the time.[22] The robustness and indeed slight strengthening of the results (see Online Appendix B.5) is consistent

[22] As explained in Online Appendix B.5, we consider varying degrees of mistake-propensity; our conclusions remain the same. We also note these checks are correlated. For example the first two samples contain a substantially smaller fraction of subjects with excessively noisy reports.

with noisy subjects if anything attenuating the effect, and shows that inattention is not driving our results. Other robustness checks include checking whether subjects mistake sampling with and without replacement.[23] Finally, the fact that retractions are more effective when they are arguably easier to interpret—e.g. when the last observation is retraction—suggests that subjects do understand the information retractions convey, but that processing this information is challenging and more so than direct evidence about the state.

7.3. Individual Heterogeneity

Underinference from retractions appears to be a robust feature within our subject pool, with our results being driven by a substantial fraction of subjects, not just a small minority. To test this, we estimate the specifications in Table 1 at the *subject*-level. We report summary statics on the subject-level estimates of the coefficient of interest (Retracted signal) in Online Appendix B.6.2. It is difficult to fully decompose the heterogeneity in these estimates into underlying population heterogeneity versus sampling noise, given the small number of belief reports per subject. However, both the mean and the median of these coefficient estimates are similar to our baseline estimates, and the estimates are strictly negative for a substantial majority of the subjects (approx. 70%).

We also examine whether retractions are more effective for subjects with higher quantitative ability, proxied for by their scores on incentivized quantitative multiple-choice questions which were asked at the end of the experiment. Expanding our main specifications to account for heterogeneity with respect to quantitative ability, we fail to find any significant effects, as reported in Online Appendix B.6.

8. CONCLUSION

Our contribution is to demonstrate and quantify retraction ineffectiveness, as distinct from biases in updating from direct information. The balls-and-urns setting we use allows us to make this comparison cleanly and precisely, in an incentivized manner, and has been used widely to establish important results on belief updating, several of which we replicate. Moreover, the abstract setting suggests that the failure of retractions can be viewed as an information-processing error, and is

[23] If sampling were without replacement, observing three draws of the same color would reveal the color of the truth ball. Less than 10% of all subjects hold extreme beliefs (close to 1 or 0) in these cases. Removing these subjects from the sample leaves results virtually unchanged.

not entirely due to some context-specific feature. This contrasts our results with past work on the continued influence effect, which typically give context-specific explanations a central role.[24] Simply put, our findings underscore that "information about information" need not be treated the same as equivalent "direct information."

In the process of illustrating that retractions are in themselves treated as less informative, we formulated a number of hypotheses relating retraction failures to a number of plausible biases. Table 7 revisits each of these hypotheses, and assesses our findings. Our analysis suggests that retraction failures are due to difficulties in reasoning particular to information about information—rendering retractions harder to interpret—and not just an expression of well-known biases. We also argued that our design is the simplest possible which still enables the desired comparisons in our two main empirical tests illustrated in Figure 2.

While our main goal in this paper was to document that retractions had a differential impact, and to determine any significant sources of variation, our results point to a number of interesting potential directions for future work. We see two as being particularly natural.

First, studying what makes information about information more complex. Our experiment was designed to highlight how errors in information processing contribute to retraction failures. Beyond the tests suggested by our theoretical framework, the richness afforded to us by variation in signal timing allowed us to speak to our proposed mechanism, without altering the fundamental nature of the task at hand. Indeed, equipped with the findings of this paper, one may elucidate richer patterns concerning cognitive noise in information about information. In particular, our results point toward the need for theoretical models of costly information processing to treat information about information differently from direct information. We consider this to be not only of independent theoretical interest, but also of practical importance in understanding how to correct misinformation and improve information transmission more generally.

Second, exploring how information-provision strategies affect belief updating from retractions. In many settings—such as in the interplay between politicians and media outlets or firms and financial auditors—the information receivers obtain results from the strategic interplay between senders and third parties that fact-check their messages (e.g. Levkun, 2021). While our results

[24] These include political information (Lewandowsky et al., 2012; Nyhan and Reifler, 2010; Nyhan, 2021; Barrera et al., 2020), financial statements (Grant et al., 2021; Tan and Tan, 2009; Tan and Koonce, 2011), and jury trials (Kassin and Sommers, 1997; Thompson et al., 1981; Fein et al., 1997). Retractions are also a quintessential part of science: The Retraction Watch Database, dedicated to tracking retractions, lists over 35,000 articles, with error and failure to replicate constituting a significant fraction of the retraction notices (Brainard and You, 2018; Fang et al., 2012).

Hypothesis	Documented (✓) or not detected (✗)
1: Subjects (a) fail to fully internalize retractions, and (b) treat retractions as less informative than equivalent new information	✓
2: Updating from retractions takes longer and increases belief variance	✓
3: Retractions of more recent evidence (a) are more effective, and (b) are faster and induce lower belief variance	✓
4: Following retractions, (a) subjects update less from new observations, and (b) updating from new observations takes longer and induces greater belief variance	✓
5: Similar belief updating biases are present in updating both from new observations and from retractions	✗†
6: Retractions are ineffective only when subjects have acted on retracted observations	✗‡

Table 7: Assessment of Main Hypotheses

Notes: See Section 3 for a more complete description of each hypothesis. † Retractions reverse the direction of the biases, resulting in under-inference and anti-confirmation bias. ‡ Null effect.

suggest receivers are particularly vulnerable to biased information provision by senders, it is unclear how belief updating depends on the fact-checking policies employed. In ongoing work we examine whether beliefs are affected by changes in which information is checked, and how. If only information of a specific kind gets checked and retracted—e.g., only articles that challenge the scientific consensus get checked, only political statements supporting specific agendas—would retractions be less effective? Additionally, if only corrections are announced—as is the case in many circumstances—would people correctly infer when retractions render unretracted evidence more reliable? We believe such results would have substantial practical value, like the ones presented here.

REFERENCES

Agranov, M., G. Lopez-Moctezuma, P. Strack, and O. Tamuz (2022): "Learning Through Imitation: An Experiment," *NBER Working Paper*. 5

Ali, S. N., M. Mihm, L. Siga, and C. Tergiman (2021): "Adverse and Advantageous Selection in the Laboratory," *American Economic Review*, 111, 2152–2178. 2

Alós-Ferrer, C., E. Fehr, and N. Netzer (2018): "Time Will Tell: Recovering Preferences When Choices Are Noisy," *Journal of Political Economy*, 129, 1828–77. 7

Ambuehl, S. and S. Li (2018): "Belief Updating and the Demand for Information," *Games and Economic Behavior*, 109, 21–39. 5

Anderson, L. R. and C. A. Holt (1997): "Information Cascades in the Laboratory," *American Economic Review*, 87, 847—-62. 1

Angrisani, M., A. Guarino, P. Jehiel, and T. Kitagawa (2021): "Information Redundancy Neglect versus Overconfidence: A Social Learning Experiment," *American Economic Journal: Microeconomics*, 13, 163–97. 7

Ba, C., J. A. Bohren, and A. Imas (2022): "Over and Underreaction to Information," *Working Paper.* 1, 2

Banks, W. P., M. Fujii, and F. Kayra-Stewart (1976): "Semantic Congruity Effects in Comparative Judgments of Magnitudes of Digits," *Journal of Experimental Psychology: Human Perception and Performance*, 2, 435–447. 8

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020): "Fake news, fact-checking and information in times of post-truth politics," *Journal of Public Economics*, 182. 38

Benjamin, D. (2019): "Errors in Probabilistic Reasoning and Judgment Biases," in *Handbook of Behavioral Economics*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier Press. 1, 2, 6, 19, 3

Bhui, R. and S. J. Gershman (2018): "Decision by Sampling Implements Efficient Coding of Psychoeconomic Functions," *Psychological Review*, 125, 985–1001. 7

Brainard, J. and J. You (2018): "What a massive database of retracted papers reveals about science publishing's 'death penalty'," Accessed: 2022-04-14. 38

Buckley, P. B. and C. B. Gillman (1974): "Comparison of Digits and Dot Patterns," *Journal of Experimental Psychology*, 103, 1131–1136. 8

Charness, G. and C. Dave (2017): "Confirmation bias with motivated beliefs," *Games and Economic Behavior*, 104, 1–23. 3

Charness, G., R. Oprea, and S. Yuksel (2021): "How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment." *Journal of the European Economic Association*, 19, 1656–1691. 5

Cripps, M. (2021): "Divisible Updating," *Working Paper.* 6

Ecker, U. K. H., S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen (2022): "The psychological drivers of misinformation belief and its resistance to correction," *Nature Reviews Psychology*, 1, 13–29. 4

Enke, B. and T. Graeber (2020): "Cognitive Uncertainty," *Working Paper.* 7, 15

Enke, B., T. Graeber, and R. Oprea (2023): "Complexity and Time," *Working Paper.* 2

Epstein, L. and Y. Halevy (2020): "Hard-to-Interpret Signals," *Working Paper.* 8

Esponda, I., R. Oprea, and S. Yuksel (2022): "Contrast-Biased Evaluation," *Working Paper.* 5

Esponda, I. and E. Vespa (2014): "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 6, 180–202. 2

——— (2021): "Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," *Working Paper.* 2

Fang, F. C., R. G. Steen, and A. Casadevall (2012): "Misconduct accounts for the majority of retracted scientific publications," *Proceedings of the National Academy of Sciences*, 109, 17028–17033. 38

Fein, S., A. L. McCloskey, and T. M. Tomlinson (1997): "Can the Jury Disregard that Information? The Use of Suspicion to Reduce the Prejudicial Effects of Pretrial Publicity and Inadmissible Testimony," *Personality and Social Psychology Bulletin*, 23, 1215–1226. 38

Friedman, D. (1998): "Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly," *American Economic Review*, 88, 933–946. 9

Frydman, C. and L. J. Jin (2022): "Efficient Coding and Risky Choice," *Quarterly Journal of Economics*, 137, 161–213. 3, 7, 9

Frydman, C. and S. Nunnari (2022): "Cognitive Imprecision and Strategic Behavior," *Working Paper.* 9

Fudenberg, D., P. Strack, and T. Strzalecki (2018): "Speed, Accuracy, and the Optimal Timing of Choices," *American Economic Review*, 108, 3651–84. 7

Gonçalves, D. (2022): "Sequential Sampling and Equilibrium," *Working Paper.* 7

Grant, S., F. Hodge, and S. Seto (2021): "Can Prompting Investors to be in a Deliberative Mindset Reduce Their Reliance on Fake News?" *Working Paper.* 38

Grether, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *The Quarterly Journal of Economics*, 95, 537–557. 19, 31, 2

Gul, F., P. Natenzon, and W. Pesendorfer (2021): "Random Evolving Lotteries and Intrinsic Preference for Information," *Econometrica*, 89, 2225–2259. 5

Gupta, N., L. Rigotti, and A. Wilson (2021): "The Experimenters' Dilemma: Inferential Preferences over Populations," *Working Paper.* 35

Halim, E., Y. E. Riyanto, and N. Roy (2019): "Costly Information Acquisition, Social Networks, and Asset Prices: Experimental Evidence," *Journal of Finance*, 74, 1975—-2010. 1

Hartzmark, S. M., S. D. Hirshman, and A. Imas (2021): "Ownership, Learning, and Beliefs," *Quarterly Journal of Economics*, 136, 1665–1717. 5

Hossain, T. and R. Okui (2013): "The Binarized Scoring Rule," *Review of Economic Studies*, 80, 984–1001. 15

Johnson, H. M. and C. M. Seifert (1994): "Sources of the Continued Influence Effect: When Misinformation in Memory Affects later Influences," *Journal of Experimental Psychology*, 20, 1420–1436. 4

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292. 6

Kassin, S. M. and S. R. Sommers (1997): "Inadmissible Testimony, Instructions to Disregard, and the Jury: Substantive Versus Procedural Considerations," *Personality and Social Psychology Bulletin*, 23, 1046–1054. 38

Khaw, M. W., Z. Li, and M. Woodford (2021): "Cognitive Imprecision and Small-Stakes Risk Aversion," *Review of Economic Studies*, 88, 1979–2013. 3, 7

Krajbich, I., C. Armel, and A. Rangel (2010): "Visual fixations and the computation and comparison of value in simple choice," *Nature Neuroscience*, 13, 1292–1298. 7

Krajbich, I., D. Lu, C. Camerer, and A. Rangel (2012): "The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions," *Frontiers in Psychology*, 3, 235–251. 3

Levkun, A. (2021): "Communication with Strategic Fact-checking," *Working Paper*. 38

Lewandowsky, S., U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook (2012): "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest*, 13, 106–131. 4, 38

Liang, Y. (2020): "Learning from unknown information sources," *Working Paper*. 8

Martínez-Marquina, A., M. Niederle, and E. Vespa (2019): "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 109, 3437–3474. 2

Masatlioglu, Y., A. Y. s. Orhun, and C. Raymond (2021): "Intrinsic Information Preferences and Skewness," *Working Paper*. 5

Miller, J. and A. Sanjurjo (2019): "A Bridge from Monty Hall to the Hot Hand: The Principle of Restricted Choice," *Journal of Economic Perspectives*, 33, 144–162. 9

Mobius, M. M., M. Niederle, P. Niehaus, and T. Rosenblat (2022): "Managing Self-Confidence: Theory and Experimental Evidence," *Management Science*, 68, 7793–7817. 15

Nyhan, B. (2021): "Why the backfire effect does not explain the durability of political misperceptions," *Proceedings of the National Academy of Sciences*, 118. 38

Nyhan, B. and J. Reifler (2010): "When corrections fail: The persistence of political misperceptions," *Political Behavior*, 32, 303–330. 38

Oprea, R. (2020): "What Makes a Rule Complex?" *American Economic Review*, 110, 3913–3951. 2

——— (2022): "Simplicity Equivalents," *Working Paper*. 2

Oprea, R. and S. Yuksel (2022): "Social Exchange of Motivated Beliefs," *Journal of the European Economic Association*, 20, 667–699. 5

Ratcliff, R. (1978): "A Theory of Memory Retrieval," *Psychological Review*, 85, 59. 8

Ratcliff, R., P. L. Smith, S. D. Brown, and G. McKoon (2016): "Diffusion Decision Model: Current Issues and History," *Trends in Cognitive Sciences*, 20, 260–281. 7

Shishkin, D. and P. Ortoleva (2021): "Ambiguous Information and Dilation: An Experiment," *Journal of Economic Theory*, Forthcoming. 8

Snowberg, E. and L. Yariv (2021): "Testing the Waters: Behavior across Participant Pools," *American Economic Review*, 111, 687–719. 35

Tan, H.-T. and S.-K. Tan (2009): "Investors' reactions to management disclosure corrections: Does presentation format matter?" *Contemporary Accounting Research*, 26, 605–626. 38

Tan, S.-K. and L. Koonce (2011): "Investors' reactions to retractions and corrections of management earnings forecasts," *Accounting, Organizations and Society*, 36, 382–397. 38

Thaler, M. (2021): "Overinference from Weak Signals, Underinference from Strong Signals, and the Psychophysics of Interpreting Information," *Working Paper*. 1, 7, 3

Thompson, W. C., G. T. Fong, and D. L. Rosenhan (1981): "Inadmissible evidence and juror verdicts," *Journal of Personality and Social Psychology*, 40, 453–463. 38

Wei, X.-X. and A. A. Stocker (2015): "A Bayesian Observer Model Constrained by Efficient Coding Can Explain 'Anti-Bayesian' Percepts," *Nature Neuroscience*, 18, 1509–1517. 7

Weizsäcker, G. (2010): "Do We Follow Others When We Should? A Simple Test of Rational Expectations," *American Economic Review*, 100, 2340—-60. 1

Online Appendix for Retractions

A. OMITTED PROOFS

Proof of Proposition 1

Let $f : [0,1] \to [0,1]$ be a strictly increasing function and ℓ denote the logit function $\ell(p) = \log(p/(1-p))$. From Definition 1, $(\ell \circ f^{-1})(b(\theta \mid S_t)) = (\ell \circ f^{-1})(b(\theta)) + \sum_{j=1}^{t} K(s_j)$. Now, let $\alpha(\tau \mid S_t) = \frac{P(\text{Retraction of } s_\tau \mid S_t, \theta=1)}{P(\text{Retraction of } s_\tau \mid S_t, \theta=-1)}$. With symmetric noise, if signal s_τ is retracted, the Bayesian update should be

$$P(\theta \mid S_t, n_\tau = 1) = \frac{P(\theta)K(s_\tau)^{\eta_t - s_\tau}\alpha(\tau \mid S_t)}{P(-\theta) + P(\theta)K(s_\tau)^{\eta_t - s_\tau}\alpha(\tau \mid S_t)},$$

where $\eta_t := \sum_{\ell=1}^{t} s_\ell$. For a retraction, the log-odds update of a Bayesian decisionmaker is therefore:

$$\ell\left(P(\theta \mid S_t, n_\tau = 1)\right) = \ell\left(P(\theta \mid S_t)\right) - K(s_\tau)\mathbf{1}[s_\tau \text{ retracted}] + \log(\alpha(\tau \mid S_t)). \tag{10}$$

Notice that $\alpha(\tau \mid S_t) = 1$, and hence $\log(\alpha(\tau \mid S_t)) = 0$, for all verifying retractions. Therefore, for any $\tau \in \{1, \ldots, t\}$,

$$\begin{aligned}
(\ell \circ f^{-1})(b(\theta \mid S_t, n_\tau = 1)) &= (\ell \circ f^{-1})(b(\theta \mid S_t)) - K(s_\tau) \\
&= (\ell \circ f^{-1})(b(\theta)) + \sum_{j \in \{1,\ldots,t\}} K(s_j) - K(s_\tau) \\
&= (\ell \circ f^{-1})(b(\theta)) + \sum_{j \in \{1,\ldots,t\} \setminus \tau} K(s_j) \\
&= (\ell \circ f^{-1})(b(\theta \mid S_t \setminus s_\tau)).
\end{aligned}$$

As $(\ell \circ f^{-1})$ is injective, then $b(\theta \mid S_t, n_\tau = 1) = b(\theta \mid S_t \setminus s_\tau)$.

If, moreover, $K(s_{t+1}) = -K(s_\tau)$, then it is immediate that $(\ell \circ f^{-1})(b(\theta \mid S_t, n_\tau = 1)) = (\ell \circ f^{-1})(b(\theta \mid S_t \cup s_{t+1}))$ and therefore $b(\theta \mid S_t, n_\tau = 1) = b(\theta \mid S_t \cup s_{t+1})$. □

	(1)	(2)
	l_t	l_t
Prior (l_{t-1})	0.834***	0.800***
	(0.037)	(0.037)
Signal (s_t)	1.126***	0.998***
	(0.071)	(0.072)
Signal Confirms Prior ($s_t \cdot c_t$)	–	0.417***
		(0.135)
R-Squared	0.41	0.41
Observations	18491	18491

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Updating from New Observations

Notes: This table reports updating from new ball draws. The sample consists of all elicited beliefs of subjects in the baseline treatment (beliefs elicited each period), excluding those elicited after a verification, within a given round. The outcome is the log-odds of beliefs in period t, l_t. s_t is the signal in round t (+1 or -1, multiplied a constant factor of Bayesian updating such that the coefficient on s_t would be 1 under Bayesian updating). $c_t := \mathbf{1}\{\text{sign}(l_{t-1}) = \text{sign}(s_t)\}$ denotes whether the signal at t confirms the prior at $t-1$.

B. TABLES AND FIGURES

B.1. Updating from New Observations: Details

In this section, we report standard Grether-style (Grether, 1980) log-odds regressions restricting the sample to histories in which no retraction occurs, thus enabling a direct comparison to existing experimental results on belief updating.

Specifically, Table 8 shows the following specification, restricted to the cases where there has not been a retraction (so only new observations):

$$l_t = \beta_0 + \beta_1 \cdot l_{t-1} + \beta_2 \cdot s_t \cdot K \tag{11a}$$

$$\text{and} \qquad l_t = \beta_0 + \beta_1 \cdot l_{t-1} + \beta_2 \cdot s_t \cdot K + \beta_3 \cdot s_t \cdot K \cdot c_t \tag{11b}$$

where t is the period, l_t is the log-odds of the beliefs reported at t, s_t is the signal in round t (+1 or -1), $c_t := \mathbf{1}\{\text{sign}(l_{t-1}) = \text{sign}(s_t)\}$ is an indicator function that equals 1 when the signal at t confirms the prior at $t-1$, and $K > 0$ is a constant factor of Bayesian updating (the log-likelihood of the signal).

In the above regression, a Bayesian decisionmaker would exhibit a coefficient $\beta_2 = 1$. Benjamin (2019) notes that this tends not to be the case: for the two incentivized studies with sequential observations he reviews, the estimated coefficient is $.528$. Thaler (2021) provides evidence that subjects overinfer (resp. underinfer) from signals in similar symmetric environments whenever $P(s_t = \theta \mid \theta) \geq 1/2$ is below (resp. above) approximately 3/5, coinciding with our parameters in the experimental design.

In the most parsimonious of our specifications, we $\hat{\beta}_2 = 1.126$, indicating mild over-inference from new observations, although not statistically different from 1 (p-value = .0762). Once we include the effect of confirmatory information, the coefficient becomes virtually equal to 1 (p-value = .976), while $\beta_3 > 0$ (p-value$< .001$). This overinference from confirmatory information—that is, $\beta_2 + \beta_3 > 1$ (p-value $= .0018$)—has been previously documented (e.g. Charness and Dave, 2017). Together, this suggests our subjects slightly over-react to new observations but that this is mostly driven by confirmation bias: they update more from a signal when the belief movement is in the direction of their prior. We also verify another deviation from Bayesian updating identified in the literature: subjects exhibit base-rate neglect. In other words, they underweight the prior, as evidenced by $\beta_1 < 1$.

To summarize, in our analysis of this data, we do not find any significant departures from existing literature on belief updating: while subjects depart from Bayesian updating, they do so in a way consistent with what one would expect from the literature. In Online Appendix B.5.1, we reestimate the specifications in Table 8 using probability weights so as to render different histories equally likely. Not only do the conclusions remain unchanged, the estimates are extremely similar.

B.2. Updating from Retractions: by Bayesian Posterior

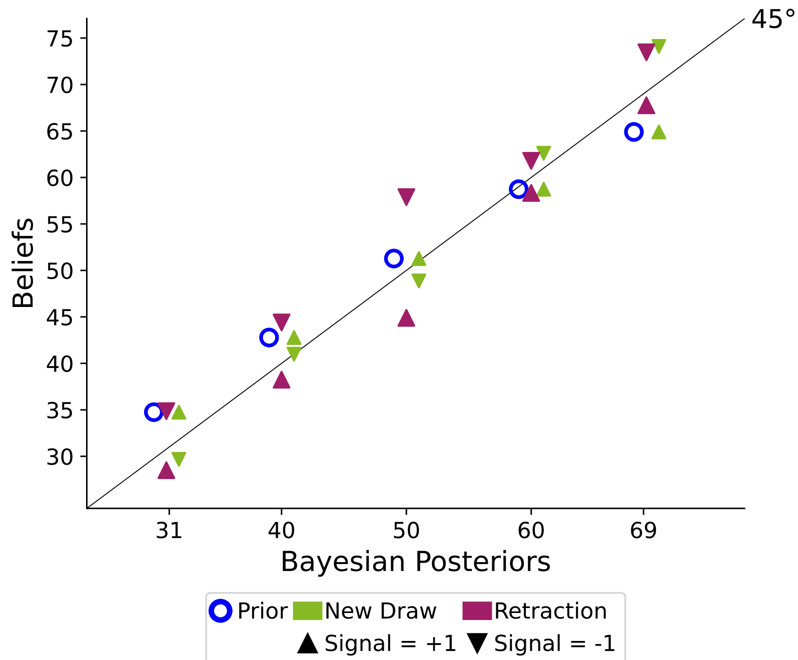


Figure 5: Retractions are Ineffective: Beliefs (Hypothesis 1)

Notes: The figure compares mean reported beliefs, disaggregated by histories conducive to the same Bayesian posterior. Blue circles denote mean reported beliefs, for period 1 and 2 histories inducing a given Bayesian posterior. Green triangles denote mean reported beliefs following a new observation, for period 3 and 4 histories inducing a given Bayesian posterior. Magenta triangles denote mean reported beliefs following a retraction, for period 3 and 4 histories inducing a given Bayesian posterior. Triangles pointing up (resp. down) indicate that the last signal was +1 (resp. -1).

B.3. Comparisons of Belief Reports to Bayesian Posteriors

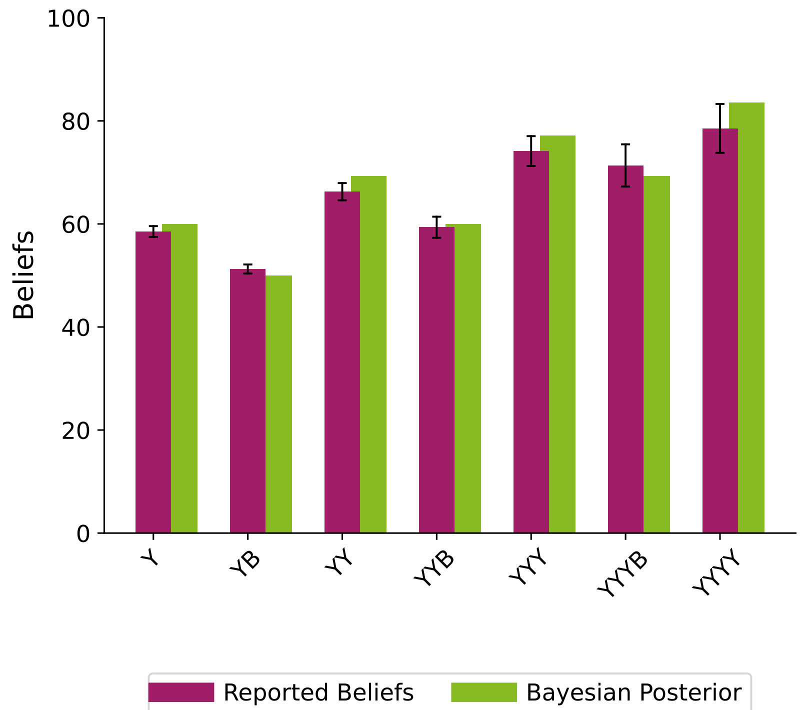


Figure 6: Reported Beliefs and Bayesian Posteriors

Notes: The figure compares mean reported beliefs with Bayesian posteriors. Histories and belief reports are symmetrized around 50, e.g. history BB is treated as YY and $b|BB$ is treated as $100 - b|YY$. The sample is restricted to the baseline treatment and to histories in which no observation is retracted. The whiskers denote 95% confidence intervals using standard errors clustered at the subject level.

(a) Difference with respect to Bayesian Posterior

(b) Distance to Bayesian Posterior

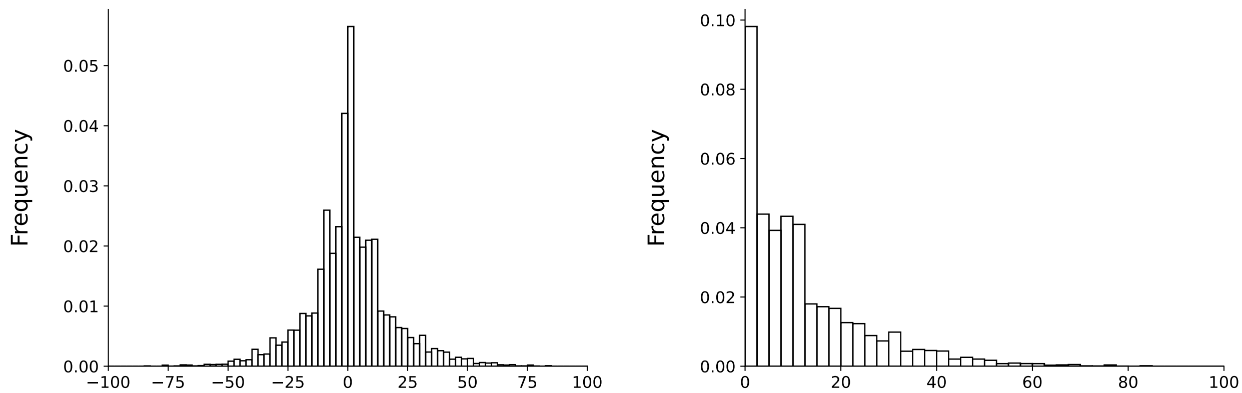


Figure 7: Distribution of Reported Beliefs

Notes: The figure shows the distribution of the the difference (a) and the absolute difference (b) of reported beliefs and Bayesian posteriors. The sample is restricted to histories in which no observation is retracted.

B.4. Sample Characteristics

	(1)	(2)	(3)
	All	Single-Elicitation	Baseline Treatment
Number of Subjects	419	204	215
Age	38.58	39.52	37.66
	(15.66)	(19.84)	(10.06)
Female	0.40	0.40	0.40
High School	0.11	0.11	0.10
College	0.21	0.23	0.19
Bachelor's or equivalent	0.45	0.41	0.49
Postgrad or equivalent	0.18	0.19	0.18
Other Education level	0.05	0.05	0.04
Answered all numeracy questions correctly	0.57	0.55	0.59
Total score on numeracy measures	1.74	1.79	1.69
	(1.01)	(1.00)	(1.02)

Table 9: Sample Characteristics

Notes: This table provides a comparison of the socio-demographic characteristics of the subjects in our sample. Column (1) considers all subjects and columns (2) and (3) provide summary statics by treatment.

B.5. Robustness Checks

B.5.1. Tables 8 and 5 with Probability Weights

In contrast to Tables 8 and 5, Tables 10 and 11 weight observations so as to make histories equally likely. We note that our conclusions remain the same.

We weight histories of observations according to the relative frequency of the associated permuted sign history within the respective period and taking into account whether or not a retraction occurred. This ensures that observations in each period are made to count equally toward the estimation, and that relative likelihood of permuted sign histories is rendered the same regardless of whether a retraction occurred or not.

	(1) l_t	(2) l_t
Prior (l_{t-1})	0.880***	0.817***
	(0.050)	(0.052)
Signal (s_t)	1.385***	1.054***
	(0.088)	(0.085)
Signal Confirms Prior ($s_t \cdot c_t$)	–	0.828***
		(0.179)
R-Squared	0.47	0.47
Observations	18491	18491

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Updating from New Draws; with frequency-weighted observations

Notes: This table reports updating from new ball draws. It re-estimates the specifications in Table 8, under the same conditions, but using inverse frequency weights.

	All Periods		Period 3	
	(1)	(2)	(3)	(4)
	l_t	l_t	l_t	l_t
Prior (l_{t-1})	0.895***	0.817***	0.949***	0.840***
	(0.054)	(0.058)	(0.064)	(0.067)
Signal (s_t)	1.552***	1.085***	2.143***	1.515***
	(0.106)	(0.103)	(0.194)	(0.250)
Signal Confirms Prior ($s_t \cdot c_t$)	–	1.083***	–	1.209***
		(0.222)		(0.389)
Retraction (r_t)	-0.040	-0.032	-0.030	-0.017
	(0.028)	(0.028)	(0.047)	(0.048)
Retraction x Prior ($r_t \cdot l_{t-1}$)	-0.020	0.068	-0.140**	-0.003
	(0.059)	(0.065)	(0.067)	(0.067)
Retracted Signal ($r_t \cdot s_t$)	-1.167***	-0.622***	-1.785***	-1.031***
	(0.126)	(0.125)	(0.210)	(0.262)
Retraction x Signal Confirms Prior ($r_t \cdot s_t \cdot c_t$)	–	-1.275***	–	-1.550***
		(0.246)		(0.384)
R-Squared	0.50	0.51	0.44	0.44
Observations	22578	22578	6081	6081

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Similar Updating Biases (Hypothesis 5); with frequency-weighted observations

Notes: This table tests whether biases in belief updating are different when updating from retractions. It re-estimates the specifications in Table 5, under the same conditions, but using inverse frequency weights.

B.5.2. Log Odds

In this section, we re-estimate the main specifications in the paper, but using log-odds beliefs as the dependent variable instead.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	l_t	l_t	Δl_t	Δl_t
Retraction (r_t)	0.005	-0.015	-0.044*	-0.043
	(0.023)	(0.034)	(0.025)	(0.029)
Retracted Signal ($r_t \cdot s_t$)	-0.233***	-0.235***	-0.249***	-0.276***
	(0.039)	(0.050)	(0.046)	(0.048)
Signal (s_t)	–	–	–	0.668***
				(0.055)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.26	0.26	0.14	0.14
Observations	22578	22578	22578	9074

Clustered standard errors at the subject level in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Retractions are Ineffective (Hypothesis 1); Log-odds

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. It re-estimates the specifications in Table 1, under the same conditions, but utilizing log-odds beliefs rather than beliefs in levels. Log-odds beliefs winsorized at 100-e and e, with $e = .1$. The results are robust to changing the winsorization threshold.

B.5.3. Comprehension Questions Correct at First Try

In this section, we re-estimate the main specifications in the paper, but restricting to subjects who correctly answered all the comprehension questions at first try.

Figure 8 shows the proportion of subjects who successfully answered the comprehension questionnaire by the n-th try and compares this with the case in which they would be choosing uniformly at random. In particular, we take the case of a sophisticated randomizer that understands which questions were incorrect and only randomizes among the ones that were not revealed incorrect.

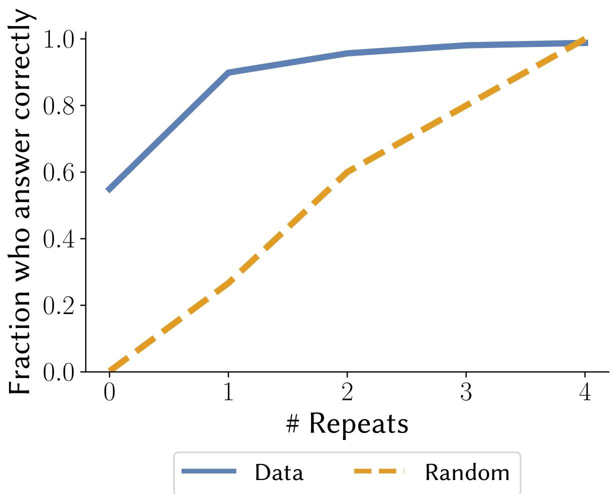


Figure 8: Comprehension Questions

Notes: The comparison is to the case in which subjects randomize uniformly over answers that were not previously tried and only in questions that were marked wrong.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	0.276	-0.003	0.071	0.162
	(0.305)	(0.331)	(0.349)	(0.362)
Retracted Signal ($r_t \cdot s_t$)	-3.522***	-4.129***	-3.779***	-3.549***
	(0.724)	(0.778)	(0.762)	(0.760)
Signal (s_t)	–	–	–	9.712***
				(0.502)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.47	0.47	0.27	0.26
Observations	13574	13574	13574	5446

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Retractions are Ineffective (Hypothesis 1); Robustness Check 1

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. It re-estimates the specifications in Table 1, under the same conditions, but restricting the used sample to subjects who correctly answered all comprehension questions at first try.

	Prior vs. Retraction		New Draw vs. Retraction	
	(1)	(2)	(3)	(4)
	log(dt_t)	V($\mathrm{b}_t \mid h_t$)	log(dt_t)	V($\mathrm{b}_t \mid h_t$)
Retraction (r_t)	0.061***	102.4***	0.111***	66.2**
	(0.018)	(18.472)	(0.017)	(25.756)
Mean Decision Time (secs)	6.049	–	6.049	–
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R-Squared	0.02	0.03	0.03	0.03
Observations	13574	1815	13574	1815

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Retractions are Harder (Hypothesis 2); Robustness Check 1

Notes: This table tests if retractions induce longer decision times and greater belief variance. It re-estimates the specifications in Table 2, under the same conditions, but restricting the used sample to subjects who correctly answered all comprehension questions at first try.

B.5.4. Correct Belief Reports when Truth Ball is Revealed

In this section, we re-estimate the main specifications in the paper, not only restricting to subjects who correctly answered all the comprehension questions at first try, but further remove from the sample subjects who failed to correctly report beliefs close to 0 or 100 when the truth ball was revealed. In particular, we remove from the sample any subject who, when state θ is revealed, failed to report beliefs $|b_t - \theta| \leq \epsilon$. We present the results for $\epsilon = .05$, but the results are robust to the choice of small ϵ.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	0.174	-0.399	-0.069	-0.010
	(0.327)	(0.351)	(0.320)	(0.306)
Retracted Signal ($r_t \cdot s_t$)	-3.647***	-3.897***	-3.671***	-3.526***
	(0.763)	(0.815)	(0.783)	(0.791)
Signal (s_t)	–	–	–	10.434***
				(0.476)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.61	0.62	0.44	0.45
Observations	10263	10263	10263	4119

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Retractions are Ineffective (Hypothesis 1); Robustness Check 2

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. It re-estimates the specifications in Table 1, under the same conditions, but excluding subjects who did not correctly answer all comprehension questions at first try or who did not correctly report beliefs when the state was revealed.

	Prior vs. Retraction		New Draw vs. Retraction	
	(1)	(2)	(3)	(4)
	$\log(dt_t)$	$V(b_t \mid h_t)$	$\log(dt_t)$	$V(b_t \mid h_t)$
Retraction (r_t)	0.065***	85.7***	0.115***	60.8**
	(0.020)	(19.579)	(0.019)	(27.202)
Mean Decision Time (secs)	5.677	–	5.677	–
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R-Squared	0.03	0.03	0.03	0.03
Observations	10263	1378	10263	1378

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Retractions are Harder (Hypothesis 2); Robustness Check 2

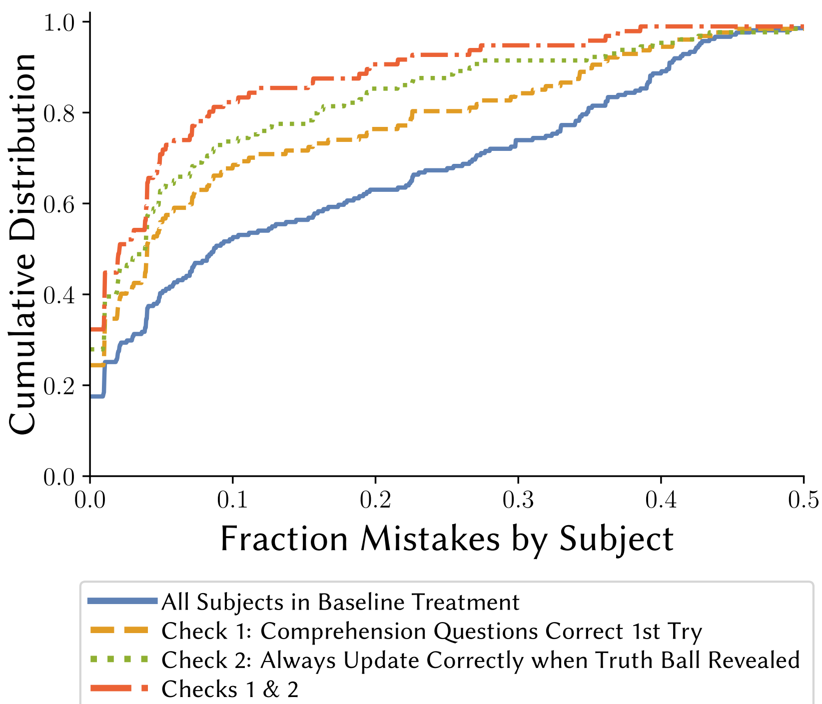
Notes: This table tests if retractions induce longer decision times and greater belief variance. It re-estimates the specifications in Table 2, under the same conditions, but excluding subjects who did not correctly answer all comprehension questions at first try or did not correctly report beliefs when the state was revealed.

B.5.5. Noisy Belief Reports

In this section, we re-estimate the main specifications in the paper, removing subjects who seem to be answering randomly.

For this purpose, we consider that the subject makes a mistake when they update their beliefs in the opposite direction to the signal, i.e. $(b_t - b_{t-1}) \cdot s_t < 0$. For each subject, we compute updating mistakes as a fraction of the total number of belief elicitations and, in Figure 9, we show the distribution of the individual-level mistakes by robustness check. It is immediate that the previous robustness checks do reduce the fraction of subjects who seem to be answering randomly.

Figure 9: Subject-Level Mistakes



In the following tables, we remove from the sample any subject who updates their beliefs in the opposite direction to the signal more than $x\%$ of the time. We present the results for $x = 10\%$; the coefficients are virtually unchanged when considering 5% and 25% instead.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	-0.586**	-0.730**	-0.692**	-0.457
	(0.286)	(0.309)	(0.283)	(0.279)
Retracted Signal ($r_t \cdot s_t$)	-5.105***	-5.912***	-5.748***	-5.360***
	(0.726)	(0.802)	(0.760)	(0.766)
Signal (s_t)	–	–	–	11.896***
				(0.404)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.69	0.70	0.50	0.46
Observations	11772	11772	11772	4732

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Retractions are Ineffective (Hypothesis 1); Robustness Check 3

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. It re-estimates the specifications in Table 1, under the same conditions, but excluding subjects who made mistakes in more than 10% of the periods.

	Prior vs. Retraction		New Draw vs. Retraction	
	(1)	(2)	(3)	(4)
	log(dt$_t$)	V(b$_t$ | h_t)	log(dt$_t$)	V(b$_t$ | h_t)
Retraction (r_t)	0.071***	122.0***	0.133***	102.2***
	(0.022)	(20.368)	(0.020)	(18.116)
Mean Decision Time (secs)	6.016	–	6.016	–
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R-Squared	0.03	0.06	0.04	0.06
Observations	11772	1589	11772	1589

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Retractions are Harder (Hypothesis 2); Robustness Check 2

Notes: This table tests if retractions induce longer decision times and greater belief variance. It re-estimates the specifications in Table 2, under the same conditions, but excluding subjects who made mistakes in more than 10% of the periods.

B.5.6. Sampling with or without Replacement

In this section, we re-estimate the main specifications in the paper, removing subjects who always report extreme beliefs (above 95 and below 5) when observing three draws of the same color.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	0.273	-0.229	-0.281	-0.153
	(0.292)	(0.395)	(0.385)	(0.392)
Retracted Signal ($r_t \cdot s_t$)	-2.722***	-3.211***	-3.214***	-2.792***
	(0.634)	(0.777)	(0.708)	(0.713)
Signal (s_t)	–	–	–	7.786***
				(0.521)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.31	0.31	0.16	0.13
Observations	20423	20423	20423	8199

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Retractions are Ineffective (Hypothesis 1); Robustness Check 4

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. It re-estimates the specifications in Table 1, under the same conditions, but excluding subjects report extreme beliefs when observing three draws of the same color.

	Prior vs. Retraction		New Draw vs. Retraction	
	(1)	(2)	(3)	(4)
	$\log(\text{dt}_t)$	$\text{V}(\text{b}_t \mid h_t)$	$\log(\text{dt}_t)$	$\text{V}(\text{b}_t \mid h_t)$
Retraction (r_t)	0.046^{***}	119.4^{***}	0.102^{***}	70.7^{***}
	(0.017)	(22.190)	(0.016)	(23.199)
Mean Decision Time (secs)	6.759	–	6.759	–
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R-Squared	0.01	0.02	0.02	0.02
Observations	20423	2740	20423	2740

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Retractions are Harder (Hypothesis 2); Robustness Check 4

Notes: This table tests if retractions induce longer decision times and greater belief variance. It re-estimates the specifications in Table 2, under the same conditions, but excluding subjects report extreme beliefs when observing three draws of the same color.

B.5.7. Decision Time Controlling Across Rounds

	Prior vs. Retraction	New Draw vs. Retraction
	(1)	(2)
	log(dt$_t$)	log(dt$_t$)
Retraction (r_t)	0.068**	0.114***
	(0.027)	(0.027)
Round	-0.013***	-0.013***
	(0.001)	(0.001)
Retraction (r_t) x Round	-0.001	-0.001
	(0.001)	(0.001)
Mean Decision Time	6.674	6.674
Compressed History FEs	Yes	No
Sign History FEs	No	Yes
R-Squared	0.05	0.06
Observations	22578	22578

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Retractions are Harder (Hypothesis 2a): Decision Time Controlling for Round

Notes: This table tests if retractions induce longer decision times. Under the same conditions, it re-estimates the specifications in Table 2 pertaining to decision time (columns (1) and (3) therein), but controlling for the round number. This serves as a robustness check to understand whether the effect of retractions on decision time fades away as subjects become more experienced.

B.5.8. Belief Variance by History

(1)	(2)	(3)	(4)
Permuted Compressed History	Retraction	Prior	Difference
B	17.54	15.34	2.21***
Y	16.91	14.20	2.71***
BB	23.51	17.41	6.10***
YB	18.30	13.76	4.54***
YY	20.59	16.31	4.27***
Permuted Sign History	Retraction	New Draw	Difference
YBB	17.54	17.14	0.41*
YYB	16.91	14.96	1.95***
YBBB	23.51	23.45	0.05
YYBB	18.30	15.20	3.09***
YYYB	20.59	21.50	-0.91

Table 22: Retractions are Harder (Hypothesis 2b): Belief Variance by History

Notes: This table tests if retractions induce greater belief variance. Each row tests for equality of variance of belief reports conditional on permuted compressed/sign histories and on whether or not a retraction was observed using Levene's test centered at the median. The columns (2) and (3) indicate the sample estimate of standard deviation of belief reports conditional on permuted compressed/sign histories and on whether a retraction occurred that period (column (2)) or not column (3). Column (4) shows the difference in sample standard deviations and indicates whether or not it is significant according to the outcome of Levene's test.

B.6. Heterogeneous Effects

B.6.1. Quantitative Ability

In this section, we test for the existence of heterogeneous treatment effects relative to the subjects quantitative ability. In the last part of our experiment, we posed three multiple-choice quantitative questions. The median number of correct answers per subject was two. We expand our specifications by interacting all the regressors with a dummy variable that equals 1 if the subject answered all questions correctly and 0 if otherwise.

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Retraction (r_t)	0.353	0.008	-0.408	-0.309
	(0.347)	(0.436)	(0.444)	(0.484)
Retracted Signal ($r_t \cdot s_t$)	-2.958***	-3.483***	-4.294***	-3.038***
	(0.779)	(0.887)	(0.795)	(0.890)
Signal (s_t)	–	–	–	7.821***
				(0.659)
All correct	-1.273***	-1.278***	-0.497***	-0.586*
	(0.406)	(0.407)	(0.170)	(0.340)
Retraction (r_t) x All correct	-0.531	-0.643	0.213	0.257
	(0.489)	(0.475)	(0.534)	(0.600)
Retracted signal ($r_t \cdot s_t$) x All correct	-0.644	-0.536	2.129**	-0.975
	(1.098)	(1.095)	(0.983)	(1.224)
Signal (s_t) x All correct	–	–	–	3.033***
				(0.868)
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes
R-Squared	0.34	0.34	0.18	0.16
Observations	22578	22578	22578	9074

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 23: Retractions are Ineffective (Hypothesis 1); Heterogeneous Effects

Notes: This table tests whether retractions return beliefs to what they would have been had the retracted observation never been observed, and compares their effectiveness relative to new direct information. We investigate the existence of heterogeneous effects with respect to quantitative ability by expanding our baseline specifications from Table 1 with interaction terms. 'All correct' denotes a dummy variable that equals 1 when the subject answered all quantitative questions correctly, and 0 if otherwise.

B.6.2. Subject-Level Estimates

	Prior vs. Retraction	New Draw vs. Retraction		
	(1)	(2)	(3)	(4)
	b_t	b_t	Δb_t	Δb_t
Mean	-3.285	-3.727	-3.892	-3.448
Median	-2.103	-2.602	-2.797	-3.236
Fraction Coeff < 0	0.68	0.71	0.73	0.72
Mean std error	2.867	3.77	4.294	4.777
Compressed History FEs	Yes	No	No	No
Sign History FEs	No	Yes	Yes	No
Lagged Sign History FEs	No	No	No	Yes

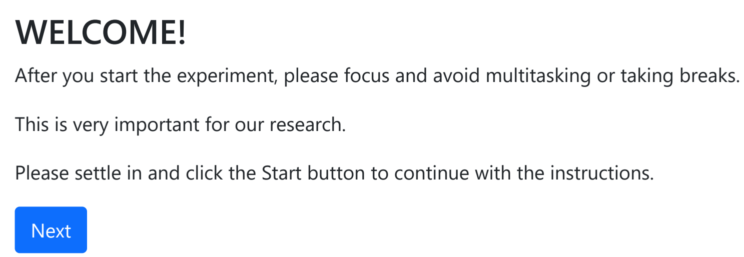
Table 24: Retractions are Ineffective (Hypothesis 1); Subject-Level Estimates

Notes: This table provides summary statistics on distribution of subject-level estimates of the coefficient of interest in the main specification of interest in this paper. We investigate the existence of individual-level heterogeneity by estimating the specifications in Table 1 for each subject. The sample consists of subjects in the baseline treatment (beliefs are elicited each period) who observed at least 8 retractions in the 32 rounds.

C. INSTRUCTIONS AND SCREENSHOTS

C.1. Start Screen and Instructions

Below are screenshots of the start screen and the instructions as presented to the subjects.



Outline

You are about to participate in an experiment on the economics of decision-making. In the experiment you can earn up to $12.50 if you do well, which will be paid to you at the end of the experiment.

You will begin, on the next screen, with the instructions. Please read them carefully.

At the end of the instructions there will be questions to check that you understand how the experiment works. Upon answering these questions correctly, you will proceed to the experiment.

The experiment contains 32 rounds, and we expect it to take **shorter than one hour** to complete. Your payment will depend on your performance in the experiment. The goal of the experiment is to study how people process new information.

Before the experiment begins there will be two practice rounds for you to familiarize yourself with the interface. After the experiment, the final part of the task is a brief survey.

You will be **guaranteed a payment of $6.00** by completing the experiment, of which $2.00 will be paid immediately afterwards and $4.00 paid together with the bonus. In addition to this, you can get a **bonus of $6.00**, which depends on your performance.

We estimate an **average hourly payment of above $9.00.**

'Bot'-Detection

This task is designed for humans and cannot be fulfilled using automated answers.

You will be asked to prove you are complying with this requirement by transcribing words at random points in this task. The text will be as legible as the text in these instructions. Any human able to read this text will be able to read the words for transcription, but a 'bot' will not. You will be allowed 3 attempts and 2 minutes per attempt. If you fail to transcribe a word three times, the task will be immediately terminated and you will automatically get no payment. You will not be able to perform the task again.

Quitting the Task

You can quit the task at any time. However, if you do so, the task is immediately terminated and you will automatically get no payment. You will not be able to perform the task again.

Additional Information

In the experiment you will answer questions which ask you to choose between different options. Your responses to this experiment will be used to study how people process information. No identifying data about you will be made available and all data we store will be anonymized. All data and published work resulting from this experiment will maintain your individual privacy.

Next

Instructions

Welcome!

In the experiment you will be asked to estimate the probability that a given ball in a box is blue or yellow.
The experiment is divided into 32 rounds, each round with up to 4 periods, plus two practice rounds before you start for you to get familiar with the interface.
We expect the overall experiment to last for less than 1 hour, although you are free to move at your own pace.
We also expect that, with an adequate amount of effort, participants get on average $9.00, of which $6.00 depends only on completing the task.

Truth Balls and Noise Balls

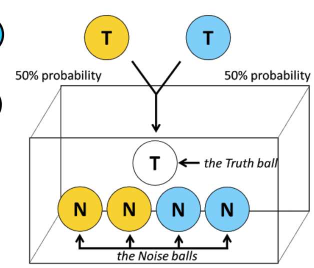
At the beginning of each round, 5 balls are put inside a box.
The balls in that box are of two kinds:

- 4 Noise Balls (N), of which 2 are yellow (N) and 2 are blue (N); and
- 1 Truth Ball (T), which can be either yellow (T) or blue (T).

Your task is to estimate the probability that the Truth Ball (T) is yellow (T) or blue (T), upon observing random draws from the selected box in each round.

Your Task

A Round



At the beginning of each round, the Truth Ball (T) is chosen to be either (T) or (T) with equal probability.

The Truth Ball (T) is then put inside the box with all 4 Noise Balls, 2 (N) and 2 (N)
.

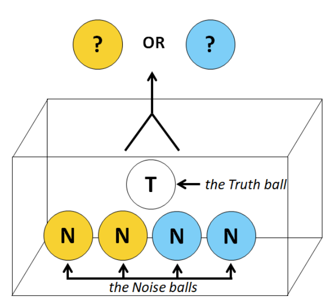
All balls remain inside the box throughout the round.

The round lasts for 4 periods, each of which may help you to guess the color of the Truth Ball (T).

Note that the Truth Ball remains the same throughout the round but changes across different rounds.

This means that the draws you observe from a particular round are not helpful to estimate the color of a Truth Ball in another round and every round you need to start afresh.

Periods 1 and 2

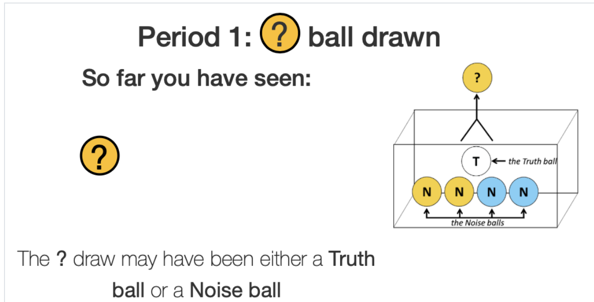


In periods 1 and 2, a ball is drawn from the box at random and you are told its color, or .
The ball is then placed back into the box.
You will not be told whether it is a Noise Ball (N) or the Truth Ball (T). Because of this, the ball will be labelled with a question mark (?).
Since the balls are drawn at random, the drawn ball (?):

- is the Truth Ball (T) with 20% probability;
- is a Noise Ball (N) with 80% probability.

Naturally, the more draws you observe, the more likely that one of them is the Truth Ball, and the more balls of one color you observe, the more likely it is that the Truth Ball is of that color. However, because in each period the ball you are shown is placed back into the box, it can be that you are shown the Truth Ball multiple times or even that you are only shown Noise Balls.

This is an example of what you can see at period 1:



Periods 3 and 4

At the beginning of period 3, a coin is flipped, and

- (i) with 50% probability it lands heads and you will observe a new draw from the box;
- (ii) with 50% probability it lands tails and you will observe a validation, learning whether one of the balls is a Noise Ball or the Truth Ball.

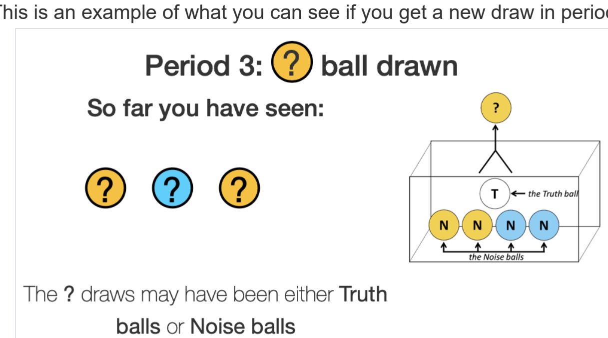
(i) New Draw

If you get a new draw, it will be exactly as before: a ball is drawn from the box and its color is shown to you, but not whether it is the Truth Ball or the Noise Ball.

Since the balls are drawn at random, the drawn ball (?):

- is the Truth Ball (T) with 20% probability;
- is a Noise Ball (N) with 80% probability.

This is an example of what you can see if you get a new draw in period 3:



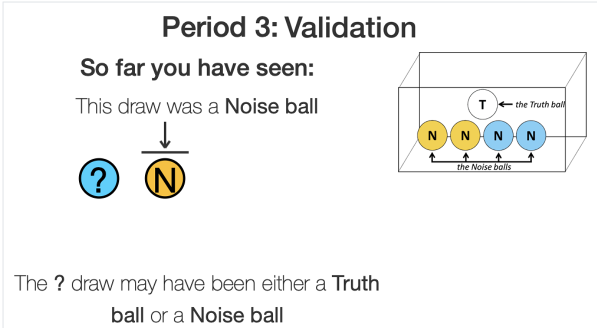
(ii) Validation

If you get a validation

one of the (?) draws is chosen at random with equal probability, regardless of whether they were draws of the Truth (T) or Noise (N) Balls.

You are then showed whether that draw was a Noise Ball (N) or the Truth Ball (T) itself.

This is an example of what you can see if you get a validation in period 3:



New Round

After these 4 periods, a new round begins.

Each round, a new color for the Truth Ball (T) is selected the same way and independently.

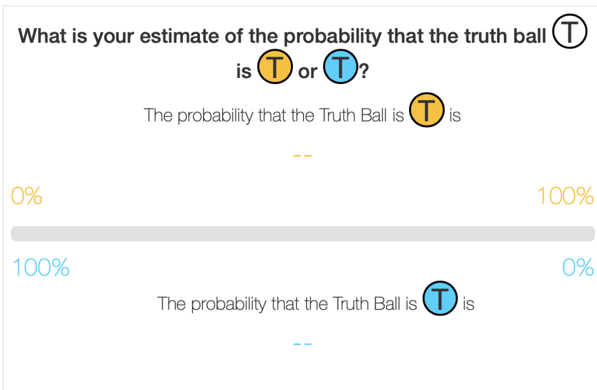
This means that whether the Truth Ball is (T) or (T) in one round has no influence on whether the Truth Ball is (T) or (T) in another round.
It will be clearly indicated when a new round begins.

Estimates

Every period and every round you will be asked to provide your estimate of the probability that the Truth Ball (T) is yellow (T) or blue (T).

Unless it is shown to you in a validation, you will not be able to know the color of the Truth Ball for sure, but you will be able to make inferences based on the draws you have seen. You will be paid based on how accurate your estimate is.

You can enter your estimate using the slider.



Payment

By completing the experiment, you can secure $6.00 for sure.
You can get a bonus of an additional $6.00 depending on your performance.

At each period, you will receive a number of points which depends on your estimate and on the color of the Truth Ball (T) in that round.

The higher the probability you assign to the correct color, the more points you get at each round.

If your estimate in a given period is that the Truth Ball is with probability q $(\times 100\%)$ and an with probability $(1-q)$ $(\times 100\%)$, then you will receive

- $100 \times (1-(1-q)^2)$ points if the Truth Ball is ; and
- $100 \times (1-q^2)$ points if the Truth Ball is .

So if your estimate completely correctly the color of the Truth Ball, you get 100 points and if you estimate completely incorrectly you get 0 points.

The lower probability you assign to the correct color, the fewer points you receive.

For instance, if you estimate that the Truth Ball is with 89% probability and with 11% probability, you receive 98.79 points if the Truth Ball is indeed and 20.79 if the Truth Ball is instead .

The points you get determine the probability of you getting the bonus.

In order to determine the probability of you getting the bonus, at the end of the experiment, one of the rounds is picked randomly with equal probability and, in this round, one of the periods is then chosen randomly, with equal probability.

The points you got = probability of getting the $6.00 bonus.
This means that if in the selected round/period you have 99.84 points you have 99.84% probability of getting the $6.00 bonus. If you have 36 points you only have 36% probability.

There is, of course, an element of chance in the task, but the more you pay attention, the more you increase the probability of getting the bonus.

All in all, the implication of the reward rule is straightforward: To maximize your expected earnings, the best thing you can do in each period is to always report your best estimate of the probability that the Truth Ball is or .

This reward system has been designed to encourage you to provide your best estimates.

Questionnaire

After you have completed all rounds, we will ask you some quantitative reasoning questions, for which you can get an extra $0.50 in bonus and then generic demographic questions.
We will not be collecting any information that allows us to identify you.
The data will be anonymized and your MTurk ID will not be available.
This data will be used for scientific research purposes only.

Only after you answer these questions will the task be completed and we will proceed to implement payments.

Questions

You must answer the following questions correctly before you can proceed.

1. There are 32 rounds and each round has up to 4 periods.

   ○ The statement is true.

   ○ The statement is false.

2. How many Noise Balls are there?

   ○ 0

   ○ 1

   ○ 2

○ 3

○ 4

3.

How many of the Noise Balls are (N) and (N)?

○ 1 (N) and 3 (N)

○ 3 (N) and 1 (N)

○ 2 (N) and 2 (N)

4.

It is possible that you see a (?) ball 3 times and the Truth Ball is (T).

○ The statement is true.

○ The statement is false.

5.

Even if in a given round the Truth Ball is (T), in the following round the Truth Ball can be either (T) or (T) with equal (50% - 50%) probability.

○ The statement is true.

○ The statement is false.

6.

If a draw you were shown (?) corresponded to a Noise Ball (N) then the Truth Ball has to be (T) and not (T).

○ The statement is true.

○ The statement is false.

7.

If a draw you were shown (?) corresponded to a Noise Ball (N) then the Truth Ball (T) may or may not be of a different color.

○ The statement is true.

○ The statement is false.

Check Answers

C.2. Practice Round

Subjects played had two practice rounds before starting the task. It was explicitly mentioned that these would not count toward their payment.

Practice Rounds

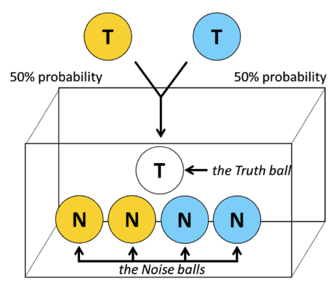
You will now play two practice rounds.
These rounds do not count towards your payment.
They are meant for you to familiarize yourself with the interface and the task.

Start Practice Rounds

Practice Round 1 of 2

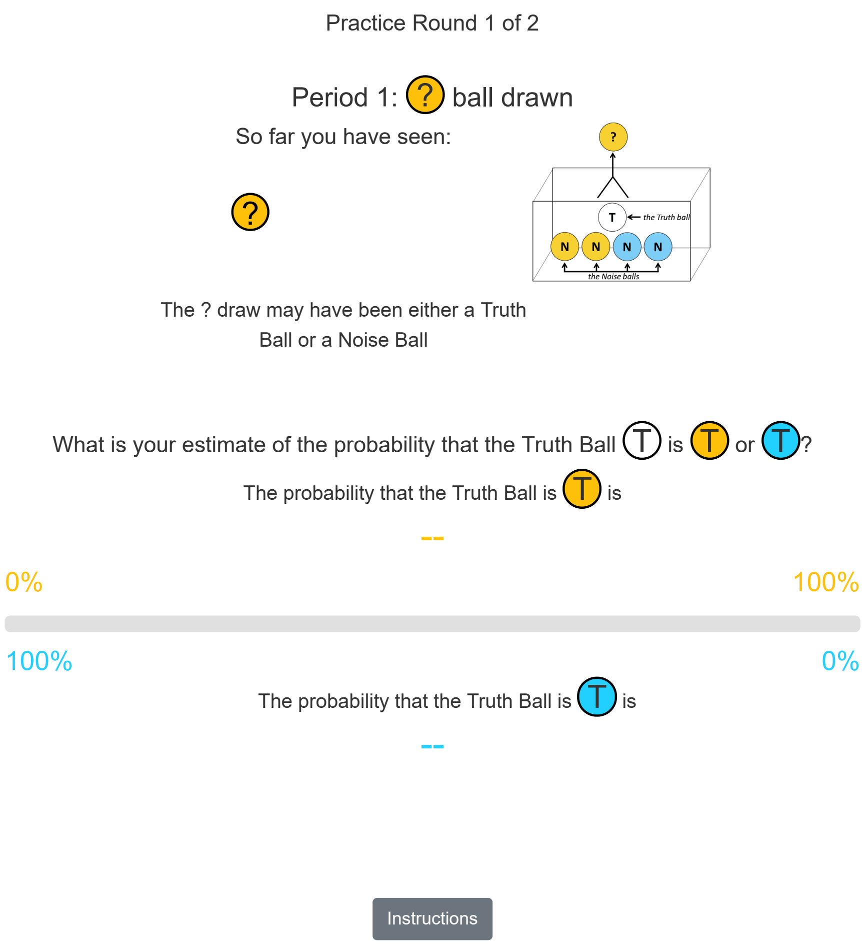
New Round

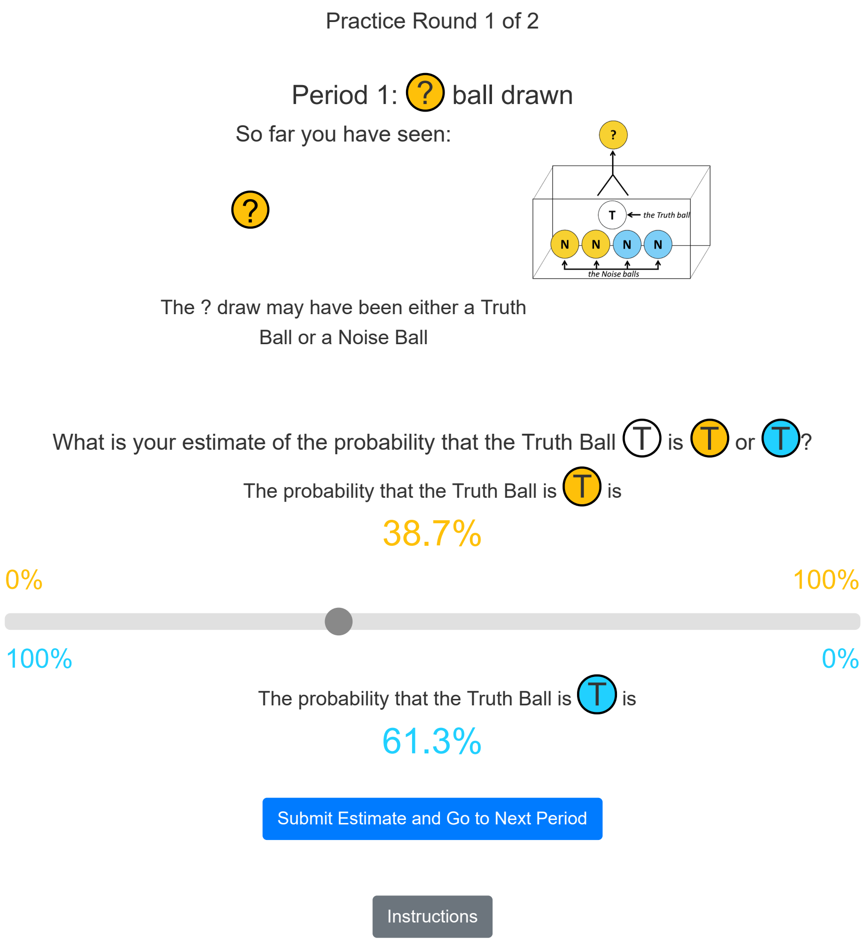
The truth ball is drawn and placed in the box



Start New Round

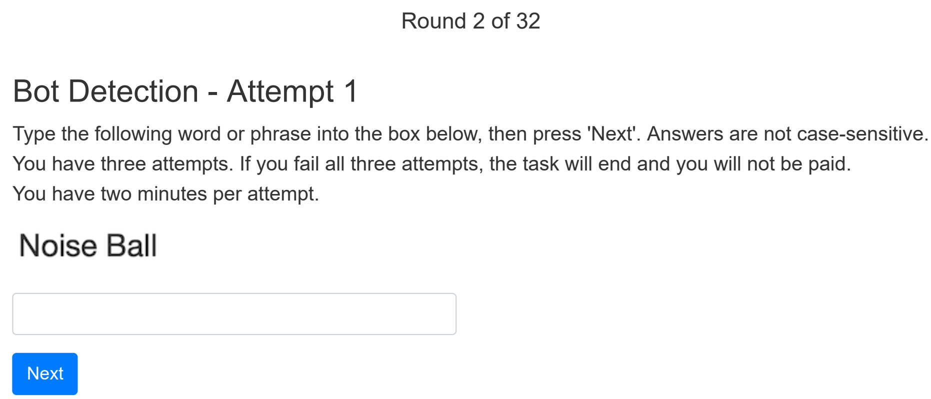
One the page loaded, the slider was blank and only activated once the subjects clicked on it.





C.3. Captchas

Subjects face five different captchas at different rounds. They had 3 tries and one minute to submit for each try. Were they to fail the 3 tries, the task ended and they would not receive any bonus.



C.4. Rounds

The rounds were described in Section 4.

Start the Task

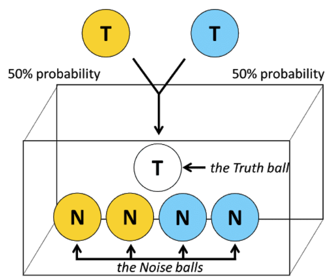
From now on, rounds matter towards your payment.

Start the Task

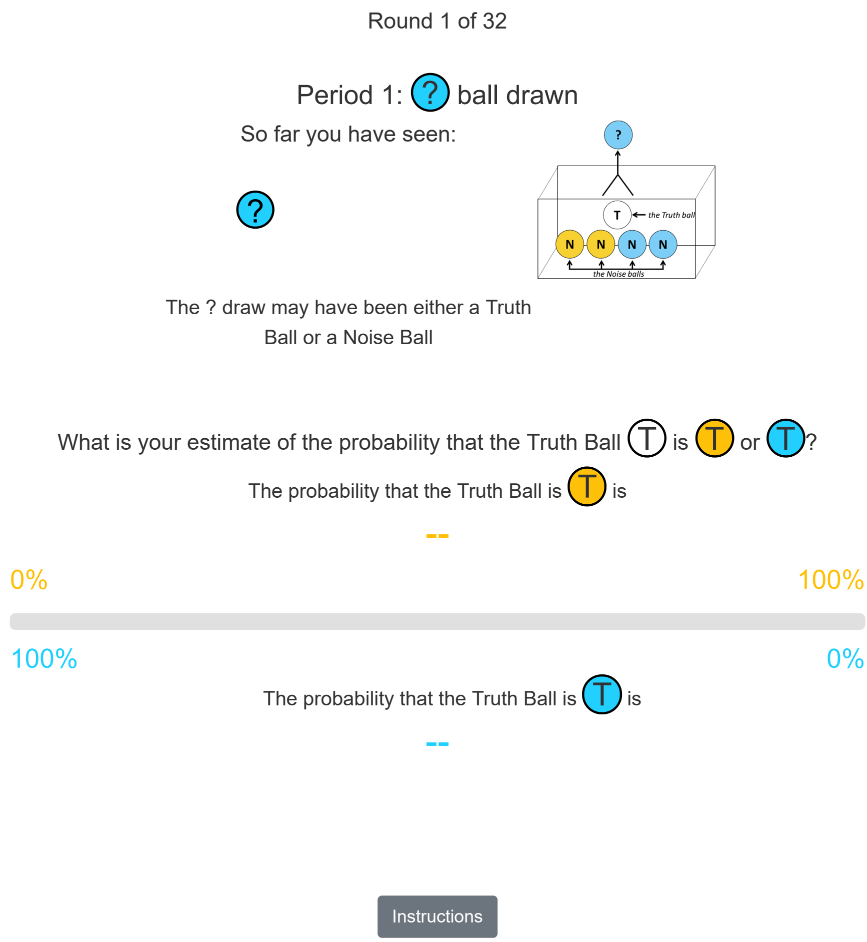
Round 1 of 32

New Round

The truth ball is drawn and placed in the box



Start New Round



C.5. Final Period Elicitation Only

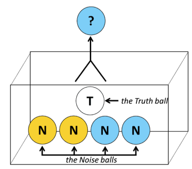
Were the subjects to be in the treatment arm in which beliefs were elicited only at the last period of each round, the last period would be just as before. In periods in which there was no belief elicitation, they would observe just the ball draw:

Round 1 of 32

Period 1: ? ball drawn

So far you have seen:

?



The ? draw may have been either a Truth Ball or a Noise Ball

Go to Next Period

Instructions

C.6. Quantitative Questions

After the main task, the subjects had to answer three questions meant to assess their quantitative ability; these were incentivized.

Questionnaire - Quantitative

In this task, you will see 3 different questions. For each, you must choose the one you believe is correct. There is only one correct answer for each. One of these 3 questions will be chosen randomly and with equal probability. If your answer to that question is correct, you will get an additional $0.50 – conditional on concluding the questionnaire and regardless of other answers or how much you have earned so far. If your answer to that question is not correct, you get no additional money.

Next

Questionnaire - Quantitative

Read each question and choose the answer that you believe is correct.

A picture was reduced on a copier to 90% of its original size and this copy was then reduced by 10%. What percentage of the size of the original picture was the final copy?

- ○ 10%
- ○ 81%
- ○ 90%
- ○ 99%
- ○ 100%

Friends Albert, Bruce and Caroline agree to buy $7 worth of lottery tickets, with Albert contributing $3, Bruce contributing $2 and Caroline contributing $2. They agree that if they win anything with any of these tickets, the winnings are to be shared out in the same ratio as their contributions. They win $175. How much does each get?

- ○ Albert gets $105, Bruce gets $35 and Caroline gets $35
- ○ Albert gets $85, Bruce gets $40 and Caroline gets $40
- ○ Albert gets $85, Bruce gets $45 and Caroline gets $45
- ○ Albert gets $75, Bruce gets $50 and Caroline gets $50
- ○ Albert gets $65, Bruce gets $55 and Caroline gets $55

In order to make 1 liter of stone paint, Navin needs to mix 3 parts (30%) of red paint, 5 parts (50%) of yellow paint and 2 parts (20%) of blue paint. If Navin has 24 liters of red paint, 40 liters of yellow paint and 6 liters of blue paint, how many liters of stone paint can Navin make?

- ○ 6 liters
- ○ 24 liters
- ○ 30 liters
- ○ 120 liters
- ○ 200 liters

Next

You must answer each question before you can continue.

C.7. Debrief and Payments

Following the task, we gathered subjects comments, socio-demographic information, and informed them of the payment they would receive.

Questionnaire - Comments

If you have any comments for the experimenters running this HIT, please leave them below. This question is optional.

Click 'Next' to complete the task.

Next

Questionnaire - Socio-Demographics

Please enter your age:

Please state your sex:

- ○ Male
- ○ Female

What is the HIGHEST LEVEL OF EDUCATION that you COMPLETED in school?

- ○ None or Primary Education: Primary School (grades 1-6)
- ○ Lower Secondary Education: Middle School or some High School incomplete
- ○ Upper Secondary Education: High School
- ○ Business, technical, or vocational school AFTER High School
- ○ Some college or university qualification, but not a Bachelor
- ○ Bachelor or equivalent
- ○ Master or Post-graduate training or professional schooling after college (e.g. law or medical school')
- ○ Ph.D or equivalent

Choose the field that best describes your PRIMARY FIELD OF EDUCATION.

- ○ Generic
- ○ Arts and Humanities
- ○ Social Sciences and Journalism
- ○ Education
- ○ Business, Administration and Law
- ○ Computer Science, Information and Communication Technologies
- ○ Natural Sciences, Mathematics and Statistics
- ○ Engineering, Manufacturing and Construction
- ○ Agriculture, Forestry, Fisheries and Veterinary
- ○ Health and Welfare
- ○ Services (Transport, Hygiene and Health, Security and Other)

Next

You must answer each question before you can continue.

Payouts

You earned $12.50.
This consists of the automatic $2.00 payment for completing the HIT, and $10.50 that will be paid to you as a bonus.

Click 'Next' to continue to the comments section. You must do this to complete the task and receive your payment.

Next

Task Complete

You have completed the HIT. Your completion code is:

9c64c7c9-cbc7-42eb-ad25-d798af4ba97f

Please copy/paste this into the space provided on the initial HIT page. You must do this in order to receive your payment.