Retractions:
Updating from Complex Information

Duarte Gonçalves[*] Jonathan Libgober[†] Jack Willis[‡]

Abstract

We modify a canonical experimental design to identify the effectiveness of retractions. Comparing beliefs after retractions to beliefs (a) without the retracted information and (b) after equivalent new information, we find that retractions result in diminished belief updating in both cases. We propose this reflects updating from retractions being more complex, and our analysis supports this: we find longer response times, lower accuracy, and higher variability. The results—robust across diverse subject groups and design variations—enhance our understanding of belief updating and offer insights into addressing misinformation.

Keywords: Belief Updating; Retractions; Information; Complexity.
JEL Classifications: D83, D91, C91.

[*] Department of Economics, University College London; duarte.goncalves@ucl.ac.uk.
[†] Department of Economics, University of Southern California; libgober@usc.edu.
[‡] Department of Economics, Columbia University; 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, Thomas Chaney, Cary Frydman, Terri Kneeland, Rani Spiegler, Michael Thaler, Séverine Toussaert, Sevgi Yuksel, and seminar audiences at Bonn, Nottingham, NYU, Caltech, Ohio State, USC, UCL, and the SWEET, and ESA conferences. Jeremy Ward, Malavika Mani, and Julen Zarate-Pina provided excellent research assistance. Funding from IEPR is gratefully acknowledged.
First posted draft: 19 June 2021. *This draft*: 13 June 2024.

1. Introduction

Retracted information often influences beliefs even once widely discredited. A notorious example is the enduring belief in a link between vaccines and autism, fueled by a subsequently retracted study in *The Lancet*. The article's impact persists as the belief in such an association remains widespread, significantly harming public health (see Pullan and Dey, 2021; Gabis et al., 2022; Motta and Stecula, 2021; Pluviano, Watt and Della Sala, 2017). While this case is illustrative, retractions are pervasive, and retracted information is rarely erased entirely.[1]

Variations of this phenomenon arise in a wide range of situations, from groundless rumors to erroneous earnings reports and from fraudulent research to false political claims. Naturally, each case is different, and retraction effectiveness in particular cases can always be attributed to unique intervening factors—e.g., special media coverage, ulterior financial or political motives, source reliability. However, while idiosyncratic factors may play important roles, issues in updating from retractions are documented too consistently and in too wide a range of settings for case-by-case explanations to be the whole story. This observation suggests moving beyond idiosyncratic factors to identify causes common to retractions generally.

In this paper, we investigate if and why there is a fundamental friction in updating beliefs from retractions. To this end, we modify a canonical experimental design to identify and quantify diminished updating from retractions relative to direct evidence, absent a variety of idiosyncratic confounds. Our analysis reveals beliefs update significantly less from retractions than from direct evidence, a finding that challenges explanations unrelated to intrinsic characteristics of retractions. We propose a simple explanation: retractions convey more complex information than direct evidence. To support this hypothesis, we present evidence based on common empirical measures of complexity—specifically, accuracy, response time, and response variability. We document these basic patterns across numerous variations, show that they are not driven by implementation details, and dismiss explanations unrelated to the processing of retractions.

Identifying the diminished effectiveness of retractions requires a clear benchmark against which updating from retractions can be measured. The aforementioned confounding factors in particular settings complicate assessing how individuals should interpret any given retraction. Whether the retraction was prompted by negligence or malfeasance, casts doubt on other evidence, is polit-

[1]Focusing on tracking retractions of academic papers, the Retraction Watch Database lists over 45,000 articles, with error and failure to replicate constituting a significant fraction of the retraction notices, in addition to misconduct (Brainard and You, 2018). Of the ten most cited retracted articles as of October 2023 in the Retraction Watch Database, seven had over a hundred citations since retraction, and two that had fewer had only been retracted in 2023 (Retraction Watch, 2023). We highlight that many papers that fail to be replicated are not retracted (Serra-Garcia and Gneezy, 2021).

ically motivated, or is disputed—all of these details influence a retraction's correct interpretation but may not be precisely quantifiable. At the same time, as individuals may err when interpreting *any* information, the mere presence of an error does not imply differential treatment of retractions compared to other pieces of new evidence. Indeed, previously documented updating biases may appear capable of explaining the diminished effectiveness of retractions. Perhaps most notable among these is *confirmation bias*—updating more when information confirms one's prior than when it does not (Rabin and Schrag, 1999)—as it suggests individuals resist disregarding information supporting their beliefs, such as a discredited study.

We develop a variation on the classic balls-and-urns experiment to identify and quantify updating from retractions absent a variety of idiosyncratic confounds. This canonical experimental design is widely used to study limitations in information processing, for example, in belief updating (Benjamin, 2019; Augenblick, Lazarus and Thaler, 2023; Ba, Bohren and Imas, 2022), social learning (Anderson and Holt, 1997; Weizsäcker, 2010; Angrisani et al., 2021), and asset pricing (Halim, Riyanto and Roy, 2019). Our version allows us to repeatedly provide retractions that are informationally equivalent to new observations, to subjects facing identical problems with access to quantifiable information about the objective truth. These properties are essential to distinguish belief updating issues specific to retractions.

We briefly describe how we modify the canonical balls-and-urns design to accommodate retractions. As is standard, subjects are presented with draws of balls from a box (with replacement), which are either blue or yellow. In our version, balls can be "noise balls," which are blue and yellow in equal proportion, or a "truth ball," which is either blue or yellow. Instead of asking if the box has a majority of blue or yellow balls, we elicit beliefs about whether the truth ball is blue or yellow, an equivalent event. After a number of draws, in which subjects are shown the color but not the truth/noise status of the ball drawn, 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 the disclosure that an earlier draw was noise as a *retraction*. In our formulation, retractions *only* provide information that a given signal was noise, analogous to a researcher having fabricated data or a news article relying on made-up claims.[2] In some practical instances, a retraction may also be coupled with additional information contradicting the initial subsequently retracted statement. Our design decouples these, as these are decoupled in several applications; however,

[2] Retractions of scientific articles are often due to problems with experimental conduct, suggesting uninformative findings but leaving open the possibility that the tested hypotheses are true. One example illustrating this possibility is the retracted study on the impact of contact on opinion formation; despite the fabrication of evidence from an early study on this topic, Broockman and Kalla (2016) subsequently conducted an experiment that did indeed provide evidence for one of its key hypotheses.

our design also allows us to study updating from retractions with additional information.

We test for retraction effectiveness by comparing beliefs over the truth ball's color after updating from retractions to (a) beliefs without having observed the retracted observation in the first place and to (b) beliefs updated from new draws with identical Bayes updates (i.e., a new draw of color opposite the retracted observation). These comparisons identify whether subjects update less from retractions than from either (a) the retracted observation or (b) a new informationally-equivalent observation. We find subjects update less from retractions in both comparisons. The magnitude of this diminished updating is significant: beliefs update on average about 50% less from retractions than new draws (see Section 3).

Why are retractions less effective? The minimality of our design strongly suggests that any explanation should be intrinsic to how retractions are processed. Consistent with this intuition, we consider a general class of *quasi-Bayesian* belief updating models that nests—but also accommodates usual deviations from—Bayesian updating. We show that results cannot be reconciled with any explanation that does not treat retractions as inherently different despite identical informational content (see Proposition 1 in Section 2.1). Notably, widely documented deviations, including confirmation bias, cannot rationalize our findings.

Our explanation is that retractions are more complex than direct information. Borrowing from Pearl (2009), we formally articulate a distinctive feature of retractions: Unlike the evidence they typically refer to, which is *directly informative* about the state—in our setting, the color of a draw—retractions are only *indirectly informative* about the state, that is, they are only informative in light of the retracted evidence.[3] Consequently, retractions always require additional contingent reasoning relative to direct evidence. Indeed, recent literature has shown not only that considering more contingencies renders problems more complex and leads to inference errors in various domains (Ali et al., 2021; Esponda and Vespa, 2014, 2021; Martínez-Marquina, Niederle and Vespa, 2019), but also that complexity considerations can explain several well-documented behavioral biases (Oprea, 2020, 2022; Ba, Bohren and Imas, 2022; Enke, Graeber and Oprea, 2023*a*). This background motivates our hypothesis that the greater complexity inherent to retractions explains diminished updating.

To test this hypothesis, we analyze three empirical complexity measures: (1) accuracy of belief reports, (2) speed of decision, and (3) variability of belief reports. All three of these data types are

[3]We say that x is directly informative about y if x and y are neither independent nor independent conditionally on some third variable z. In our setting, the color of a draw is directly informative about the state, but learning about its noise status is indirectly informative: only by conditioning on its color can it be informative, and it is otherwise independent.

borrowed from past work using them as measures of complexity and cognitive noise—see e.g., Caplin et al. (2020), and Enke and Shubatt (2023) for the first, Wright and Ayton (1988), Krajbich et al. (2012), and Frydman and Jin (2022) for the second, and Khaw, Li and Woodford (2021) and Enke and Graeber (2021) for the third. All proxies are larger when updating from retractions compared to equivalent new information, as well as compared to when the retracted signal had never been seen. These patterns suggest higher complexity for retractions, as proposed.

We leverage natural variation provided by our design, which suggests variation in the relative complexity of retractions, and verify that these co-vary with updating strength. First, we compare updating from retractions of more or less recent observations. If the most recent observation is retracted, subjects can simply "go back" to a past belief, making updating easier. This point suggests that retractions of more recent evidence are less complex, corroborated by our empirical complexity measures. In line with our mechanism, subjects also update more when the most recent observation is retracted than when retractions refer to an earlier observation. Second, we examine updating from new observations after retractions. Beliefs respond less to new observations after retractions, and our empirical measures indicate inference is more complex.[4]

We further examine how updating patterns vary across histories. We use standard Grether (1980) log-odds regressions to compare biases from retractions to those typically documented in updating from new observations. While updating from new evidence exhibits confirmation bias, retractions entail both under-inference and anti-confirmation bias. In line with this, confirmatory retractions are least effective (relative to equivalent new evidence) at histories inducing more extreme beliefs. These findings offer valuable insights into the unique influence of retractions on belief-updating behavior.

We conducted a wide range of robustness checks to ensure the validity and generalizability of our results. Firstly, we assessed whether our results simply reflect limited subject understanding and inattention despite our screening measures and attention checks. We consider removing subjects who are 'noisy' or prone to mistakes, as well as those who did not correctly answer unincentivized comprehension questions on the first try. We can also rule out misinterpreting that the draws are with replacement. A theme that emerges is that our results are maintained, if not strengthened, when restricting to subjects who appear to have understood the task better.[5]

[4]This finding is relevant for situations where (i) some evidence is found inaccurate and (ii) further contradictory evidence is subsequently revealed. The diminished updating from retractions under (i) and the diminished updating *following* retractions in (ii) indicate that both elements contribute to a diminished updating from retractions.

[5]This finding is perhaps unsurprising since documenting any effect requires that subjects act differently for retractions; if subjects answered randomly or always answered 50-50, we would not document any difference. But 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 the richness it contains.

Secondly, we explored variations in subject characteristics. We find that our results on the diminished updating from retractions and its greater complexity are robust to whether subjects perform better or worse in quantitative tasks, are more or less confident about their belief updating, are more or less experienced with the task, or more or less Bayesian in updating from observations. Although we are not powered for a fully-fledged within-subject analysis, inspection of individual heterogeneity in our results indicates that the diminished effectiveness of retractions compared with new observations is a general phenomenon in our sample.

Thirdly, we examined the impact of design variations, such as having shorter histories, omitting the history of past draws, garbling information so that the state is never perfectly learned, and different wording for retractions. We found that our results remained robust across all these different experimental setups. Notably, our results are robust even when beliefs are only elicited at the end of each round—dispelling concerns that 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. Overall, our comprehensive analysis underscores the robustness and reliability of our findings across various conditions and contexts.

These observations support the claim that our work provides some of the first evidence that diminished retraction effectiveness could have origins (at least partially) in fundamental information processing properties. An advantage of showing this in a setting where beliefs can be elicited directly is that it suggests a unified and systematic approach to analyzing patterns in belief updating from retractions. Of course, retractions in practice will differ from those we present to subjects in our experiment. Indeed, we expect many elements deliberately precluded by design, such as memory frictions, salience, or motivated reasoning, to play an significant role in many settings where retractions appear less effective.

Our results are both of practical value and theoretical interest. We designed the experiment to connect the diminished effectiveness of retractions to information processing errors.[6] From a theoretical standpoint, our findings motivate the development of theoretical models of costly information processing that treat indirect information differently from direct information—even when their informational content is the same. From a practical standpoint, our analysis provides guidelines regarding how individuals respond to retractions, potentially relevant to campaigns targeting misinformation. The fact that retraction failures arise due to information processing errors suggests limits to the "this time is different" logic policy-makers may adopt—it is generally

[6]In this sense, our paper is part of a sizable literature which, while motivated by anecdotal or domain-specific evidence of biases, uses fundamental 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).

unreasonable to expect a retraction to be entirely successful in correcting beliefs. In many real-world cases, appreciating the inability to correct beliefs with retractions ex-post may very well have changed the calculus regarding decisions to disseminate information ex-ante.[7]

1.1. Past Work on Causes and Consequences of Continued Influence

The closest precedent for the diminished effect of retractions comes from the literature on the *continued influence effect* in psychology. Reviewing this literature, Ecker et al. (2022) define this effect as the finding that "misinformation can often continue to influence people's thinking even after they receive a correction and accept it as true." Johnson and Seifert (1994) provided an early articulation of such a result, asking subjects to recount the cause of the start of a fire and finding that they would still rely upon discredited information.[8] Chan et al. (2017) and Walter and Tukachinsky (2020) provide meta-analyses of the literature—across experiments that range from stories to advertising, scientific retractions, and beyond—and find that corrections fail to fully correct beliefs. These and similar patterns have been extensively documented in many settings; Appendix A discusses specific applications.

Ecker et al. (2022) and Lewandowsky et al. (2012) discuss a number of channels for continued influence to emerge. These include biases related to memory storage (e.g., in terms of "mental models" individuals used) and retrieval,[9] as well as explanations based on the perceived credibility of a retraction and the extent to which it clashes with an individual's worldview.[10] A confounding factor, however, is that in many existing papers, the "continued influence effect" and related 'biases' could actually be consistent with Bayesian updating, depending on the implementation of retractions (see Guay et al., 2023; Pennycook et al., 2021).

Our implementation of retractions within a balls-and-urns design differs from the existing literature in that we, as analysts, know a retraction's objective informational content. This advantage facilitates the identification of differences in information processing *due to information being a retraction*, and our proposed mechanism is intrinsically tied to how retractions generate infor-

[7] 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.

[8] A more extreme reaction is *backfiring*, in which subjects believe more strongly in the retracted information. Nyhan and Reifler (2010) documented this pattern when providing subjects with information about the presence of weapons of mass destruction in Iraq during the early 2000s (and subsequently providing corrections). But unlike continued influence, backfiring has not been replicated for the most part. See Nyhan (2021) for an authoritative discussion.

[9] In particular, the mere passage of time may affect the perception of evidence (Jacoby et al., 1989).

[10] As illustrated by Susmann and Wegener (2022), a possible reason underlying this belief-updating pattern is that it reflects an implied cognitive dissonance (Harmon-Jones and Mills, 2019), owing to the psychological discomfort following from holding two contradicting ideas that retractions induce.

mation. Further, while each explanation above is undoubtedly important in some circumstances and less relevant in others, our design allows us to differentiate our proposed mechanism from these setting-specific explanations—issues discussed in more detail in Section 6.

1.2. Other Work on Belief Updating Biases

Our paper builds on the experimental literature studying belief updating. Benjamin (2019) provides a comprehensive survey; of independent interest, we replicate many of its key findings.[11]

We aim to identify and distinguish the updating from retractions and other well-known biases. For instance, we document *base-rate neglect* (whereby agents underweight the prior when updating; see, e.g., Esponda, Vespa and Yuksel 2024), as well as *confirmation bias*, discussed above (see also Rabin and Schrag, 1999).[12] We show in our theoretical framework that the diminished effectiveness of retractions is *distinct from these biases* and cannot be explained by models that do not treat retractions inherently differently.

Our analysis suggests that "indirect information" is more complex to process than "direct information." Though our focus on retracting information is new, the idea that contingent reasoning entails higher cognitive effort has been illustrated in different settings. One of the first documented difficulties of contingent reasoning was Charness and Levin (2005) for the winner's curse.[13] Closer to our study is Enke (2020), which documents in a pure prediction setting that many subjects consistently fail to account for the informational content from the absence of observations, suggesting a failure of contingent reasoning. One microfoundation driving greater complexity for "indirect information" than "direct information" is that subjects face *higher cognitive imprecision* in their understanding of the informativeness of a retraction than of an observation—see Woodford (2020) for a survey, and Enke and Graeber (2022) and Augenblick, Lazarus and Thaler (2023) for recent applications to belief updating.

2. Framework and Design

2.1. Information Arrival: Draws and Retractions

Our experiments consider a simple belief updating problem. Subjects form beliefs over a state θ, which takes one of two values with equal probability, say $\theta \in \{yellow, blue\}$. We write $\hat{p}_t$ to

[11] For recent papers studying these patterns in belief updating, see, for instance, Ambuehl and Li (2018), Coutts (2019), and Augenblick, Lazarus and Thaler (2023).

[12] To avoid confounding factors, our design features exogenous information; Charness, Oprea and Yuksel (2021) study how biases may influence subjects' *choice* of sources of information.

[13] See Esponda and Vespa (2014) and Martínez-Marquina, Niederle and Vespa (2019) for more on difficulties in contingent reasoning in particular games.

denote a subject's belief that $\theta = \textit{yellow}$, given all the information observed by period t. We use the term "signal" as a generic term for information throughout. Our interest is in two kinds of information subjects may have access to: *draws* and *retractions*.

Draws. In a given period t, subject i may observe $s_t \in \{\textit{yellow}, \textit{blue}\}$, a signal informative about θ and drawn independently conditional on θ. We refer to this kind of information as an "observation" or "draw." In our baseline experiment, each observation s_t can correspond either to the *truth*, in which case $s_t = \theta$, or to *noise*, in which case it is given by an independent $\epsilon_t \in \{\textit{yellow}, \textit{blue}\}$. Denoting the former event by $\{n_t = T\}$ and the latter by $\{n_t = N\}$, we focus on cases where these events are independent of θ. To summarize, we have $s_t = \theta$ if $n_t = T$, and $s_t = \epsilon_t$ if $n_t = N$, where $n_t \in \{T, N\}$ and n_t, ϵ_t and θ are independent. For simplicity, we write $S_t = \{s_1, \ldots, s_t\}$. In this setup, if $n_t = T$, then s_t reveals the state. In one variant, we additionally allow s_t to be *im*perfectly informative even when $n_t = T$, but we defer our discussion of this possibility.

Retractions. The second kind of signal a subject may receive in period t is a *retraction*. Formally:

Definition 1. *A* retraction *of the ρ-th observation informs that it was noise, i.e., $n_\rho = N$.*

Retractions provide information about *past signals*. The process by which retractions are determined—e.g., how observation ρ is chosen to be retracted—matters for how they should be interpreted, a theme we return to later. Important for identification in our experimental paradigm, we focus on the following type of retraction:

Definition 2. *A* verifying retraction *of the ρ-th observation is a retraction in which ρ (the period that the retraction refers to) is chosen independently of that or other observations' truth value.*

Our experiment implements verifying retractions by selecting ρ uniformly at random from all past observations and subsequently revealing n_ρ, that is, whether this observation was noise.[14] We refer to the signal that informs the subject of n_ρ as a *verification*, noting that a verification is a retraction when $n_\rho = N$. The indicator variable r denotes the occurrence of a retraction, whereby $r_t = 1$ if a retraction occurs in period t and $r_t = 0$ otherwise.

2.2. Experimental Design

We turn to how we operationalized this information arrival process in our experiments. Here we focus on our baseline setup, and subsequently discuss how we modified it in our variants.

In each *round* of the experiment, we provided information about a state across up to four *periods*:

[14] This implementation implies one learns that past information was not noise when $n_\rho = T$, which, in the current setting, perfectly reveals θ in turn.

1. At the start of the round, a *truth ball* (corresponding to the state θ) is chosen at random to be either yellow or blue, with equal probability. The truth ball is then placed into a box with four *noise balls*, two yellow and two blue (corresponding to $P(n_t = N) = 4/5$ and $P(\epsilon_t = yellow) = 1/2$).
2. In periods one and two, subjects obtain a *new observation*: a draw from the box with replacement. They see the ball's color (s_t) but not whether it is the truth ball or a noise ball (n_t).
3. In periods three and four, and independently across periods, subjects either obtain a new observation (as above) or observe a verification of an earlier observation (ρ) from the same round, with equal probability. Under a verification, one of the previous draws is chosen uniformly at random, and it is revealed whether that draw was a noise ball ($n_\rho = N$)—a *retraction*—or the truth ball ($n_\rho = T$). If the draw turns out to have been the truth ball, the round ends, as at that point, the state (the color of the truth ball) is fully revealed.

Subjects report their belief regarding the probability that the truth ball is blue or yellow ($\hat{p}_t$) at the end of each period, that is, after each new signal (observation or retraction). These reports are incentivized, as detailed in Section 2.3. Each subject plays a total of 32 rounds, and no feedback on performance is provided until performance-based payouts are made at the end of the experiment.

Variants. In total, we ran four experiments with six main, across-subject treatments (including the baseline). Table 1 summarizes these treatments and details where the paper discusses them. These variants aimed to demonstrate the robustness of our findings and to investigate the underlying mechanisms. Table 5 presents sample characteristics for each treatment.

Experiment	Treatment	Venue	# Subjects	Duration	Payment	Sections
A	Baseline	MTurk	211	31 min	\$11.96	Throughout
A	Elicit at End	MTurk	204	24 min	\$8.14	6.3
B	Garbled Information	MTurk	164	40 min	\$11.03	6.3
C	Baseline	Prolific	155	49 min	\$11.64	Throughout
C	Retraction Information	Prolific	164	52 min	\$11.76	6.1
C	No History	Prolific	164	51 min	\$11.80	6.3
D	Short Histories	Prolific	150	26 min	\$12.02	6.3

Table 1: Summary of Treatments

Notes: This table summarizes the four experiments, their respective treatments, and the sections of the paper where they are discussed. "Duration" and "Payment" refer to the average time spent in the experiment in minutes and to the average payment in USD, respectively.

2.3. Implementation Details

Experimental Interface. Figure 1 summarizes the explanatory visuals shown to subjects, and Online Appendix H contains the experiment's instructions. Subjects reported beliefs using a slider, which displayed the stated probability assigned to the truth ball being yellow and the complementary probability assigned to it being blue. After the instructions, subjects were given two rounds of unincentivized "practice" to familiarize themselves with the interface.

Subject Pool and Comprehension Checks. We ran four experiments (labelled A-D) comprising different treatments as described in Table 1. The first (A and B) were on Amazon Mechanical Turk (MTurk), and the remaining two (C and D) were on Prolific. We recruited 1,212 subjects in total; Appendix B presents sample characteristics. The assignment of subjects to treatments was randomized within each experiment. We took several steps to ensure that our subject pool was of high quality; Section 6.1 describes these steps in greater detail.

Payments. We incentivized subjects to report their beliefs truthfully using a binarized scoring rule (Hossain and Okui, 2013; Mobius et al., 2022). By reporting $\hat{p}_t \in [0,1]$, a subject would receive $\$High$ with probability $(1-(\mathbf{1}\{\theta = yellow\}-\hat{p}_t)^2)$ and $\$Low$ with complementary probability, where $\$High$ and $\$Low$ correspond to \$12.00 and \$6.00 for experiments A and B (ran in 2020 and 2021), and to \$13.00 and \$7.00 for experiments C and D (ran in 2024). To determine payments, we used a report from a single randomly selected period of a randomly selected round.[15]

In the instructions—but not in the main interface—we provided information on the elicitation procedure, phrased as eliciting the probability that the truth ball was either yellow or blue. The instructions explained that the procedure was meant to ensure they were incentivized to answer truthfully. As the elicitation scheme we used may appear complicated, we sought to limit the extent to which subjects were required to focus on it while maintaining transparency. Danz, Vesterlund and Wilson (2022) show that the binarized scoring rule can introduce noise and "pull beliefs toward the center," although the magnitude appears to vary across subject pools and might be lower for online platforms (see Healy and Kagel, 2023). As our primary focus is on how updating from retractions compares to direct information, any difference is still meaningful. Moreover, any potential under-reaction would make it *harder* to detect an effect of retractions, not easier.

We also asked additional questions on mathematical ability, which were incentivized via a \$0.50 reward if a randomly chosen question was answered correctly. The average duration and com-

[15] Azrieli, Chambers and Healy (2018) show that random selection is essentially the unique problem-selection mechanism inducing incentive compatibility when preferences satisfy state-wise monotonicity, namely that subjects prefer higher payments given any realization of uncertainty (selected problem/underlying states).

pensation were 31 minutes and \$9.98 (\$24.36/hour) for experiments A and B and 45 minutes and \$11.81 (\$20.52/hour) for C and D. For comparison, this rate is higher than the MTurk experiment of Enke and Graeber (2022), 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 for experiments A and B were pre-registered, as were many of our hypotheses. The hypotheses pertaining to decision time, belief variance, and belief bias and their variations were introduced subsequently, as feedback we received convinced us they provided evidence for our proposed mechanism.

3. Diminished Updating from Retractions

3.1. Theoretical Predictions

We start by clarifying how our design enables us to identify if and how updating differs between retractions and direct information. The core of our identification strategy comes from our result that, in our setting, any difference in updating would be inconsistent with any explanation that does not treat retractions differently from direct information—including the general class of frameworks used to explain many known deviations from Bayesian updating. In the process, we clarify why seemingly similar paradigms fail to do so, and the extent to which continued influence could be consistent with rational belief updating.

Let $P(\cdot)$ denote *objective* probabilities associated with the data generating process, and $\hat{P}(\cdot)$ denote i's *subjective* beliefs. For a Bayesian, subjective beliefs about θ, $\hat{p}_t := \hat{P}(\theta \mid \mathcal{H}_t)$, coincide with the objective probability that $p_t := P(\theta \mid \mathcal{H}_t)$, where $\mathcal{H}_t$ represents the entire history at period t, that is, the set of all the draws observed as well as any retractions, fixing the order. Past work has routinely rejected this hypothesis. A common alternative is to assume there is a strictly increasing f_i such that $\hat{p}_t = f_i(p_t)$. It follows that, upon observing some event E at t, updating of beliefs $\hat{p}_{t-1}$ is given by the following identity:

$$\mathscr{L}(f_i^{-1}(\hat{p}_t)) = \mathscr{L}(f_i^{-1}(\hat{p}_{t-1})) + K(E), \tag{1}$$

where $\mathscr{L}(p) := \ln\left(\frac{p}{1-p}\right)$ denotes the log-odds of $p \in (0,1)$, and $K(E) := \ln\left(\frac{P(E|\theta=yellow)}{P(E|\theta=blue)}\right)$ the log-likelihood of E, with the understanding that $\mathcal{H}_t = \mathcal{H}_{t-1} \cup E$. As long as $\hat{p}_t = f_i(p_t)$, this relationship holds for all histories $\mathcal{H}_t$; this point will be useful in our analysis.

Inspired by Cripps's (2021) axiomatic work, we call a decision-maker who updates according to (1) "quasi-Bayesian:"

Definition 3. *We say that a decision-maker is* quasi-Bayesian *if there exists a strictly increasing*

f_i such that, for any information $\mathcal{H}_{t-1}$ and event E, $\hat{p}_t = \hat{P}(\theta \mid \mathcal{H}_{t-1}, E)$ can be derived from $\hat{p}_{t-1} = \hat{P}(\theta \mid \mathcal{H}_{t-1})$ according to (1).

Note that, to accommodate some forms of confirmation bias, it may be necessary to allow the function f_i to depend on the prior belief from which subjects update; we strive to be as agnostic as possible and our comparisons will hold across a number of possible assumptions. We return to a discussion of possible microfoundations for distortions under quasi-Bayesianism in our discussion of mechanisms, in Section 4.

Our main comparisons in the paper relate to the following subjective beliefs:

(1) $\hat{P}(\theta|S_t, n_\rho = N)$: the subject's belief after observing the retraction $n_\rho = N$ in period t;
(2) $\hat{P}(\theta|S_t \setminus s_\rho)$: the subject's belief had the retracted observation s_ρ never been observed; and
(3) $\hat{P}(\theta|S_t \cup s_{t+1})$: the subject's belief following a new observation s_t instead of the retraction.

Proposition 1. *Suppose retractions are verifying. For any quasi-Bayesian,*

(a) their belief after observing the retraction $n_\rho = N$ in period t is the same as their belief had the retracted observation s_ρ never been observed, i.e., $\hat{P}(\theta|S_t, n_\rho = N) = \hat{P}(\theta|S_t \setminus s_\rho)$;
(b) their belief after observing the retraction $n_\rho = N$ in period t is the same as their belief following a new draw s_t instead of the retraction, i.e., $\hat{P}(\theta|S_t, n_\rho = N) = \hat{P}(\theta|S_t \cup s_{t+1})$, if and only if the log-likelihood of the new draw is negative of the retracted observation, $K(s_{t+1}) = -K(s_\rho)$.

The proof of this proposition essentially follows from an application of Bayes rule and the observation that quasi-Bayesian updating rules still satisfy this identity under the transformation f_i^{-1}. An identical argument could be used to introduce additional history dependence into the updating rule; our identification strategy below would remain valid. More generally, while our framework allows decision-makers to exhibit 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.

We focus on verifying retractions to ensure equivalence to signal histories with only new draws and that updating is equivalent to simply never having observed the retracted evidence, and nothing more. In particular, the log-likelihood of retracting an observation exactly offsets the log-likelihood of the retracted observation, i.e., $K(n_\rho = N) = -K(s_\rho)$. This property contrasts with setups where subjects consider restricted information structures, a factor Miller and Sanjurjo (2019) argue leads to mistakes in probabilistic reasoning, such as those in the Monty Hall Problem.[16] Here, the *selection* of a signal is independent of its and other observations' truth value,

[16] In the Monty Hall Problem, a subject selects one of three doors, one of which hides a prize. After making a choice, an *un*selected door that does *not* hide the prize is opened. The subject can then switch choices. Since only unselected doors *without a prize* are opened, the other unselected door is more likely to hide a prize, making switching optimal.

making our implementation of retractions *unrestricted*. In fact, Proposition 1 no longer generally holds if retractions are not verifying and unrestricted.

A provocative implication of this observation is that sometimes "continued influence" or related 'biases' could simply reflect Bayesian updating (Pennycook et al., 2021). If, for instance, only uninformative signals are selected ($\rho = t$ implies $n_t = N$), retracting an observation gives more credence to *non-retracted* evidence, which can lead to updating patterns resembling "continued influence" (Johnson and Seifert, 1994)—as well as patterns resembling backfiring (discussed in Nyhan, 2021).[17] While in many important settings, disclosure is targeted and retractions are restricted, verifying retractions allow direct comparisons and serve as a natural starting point—thus implying that a (quasi-)Bayesian agent would *not* exhibit continued influence.[18]

3.2. Hypothesis and Identification

This paper aims to study updating from retractions and, in particular, compare it to updating from direct information. Our first hypothesis concerns our two basic approaches to doing so:

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

We emphasize that our usage of "retractions" reflects the meaning in Definition 1, with "otherwise equivalent" reflecting the last case of Proposition 1. In our experimental setting, the log-likelihood of a *blue* draw exactly offsets that of a *yellow* draw ($K(blue) = -K(yellow)$), so a retraction of a *blue* draw is informationally equivalent to a new *yellow* draw, and vice versa.

We will refer to retractions having *diminished effectiveness* as the finding that belief updates are diminished when generated by retractions, reflecting either part of this hypothesis. Proposition 1 shows retractions should be as effective as new direct information unless subjects treat these two types of information differently. Therefore, we identify the diminished effectiveness of retractions as a phenomenon distinct from belief-updating biases that are not intrinsically related to retractions.

In our context, parts (a) and (b) of Hypothesis 1 correspond to the following comparisons, which we will make repeatedly in the paper, explained visually in Figure 1:

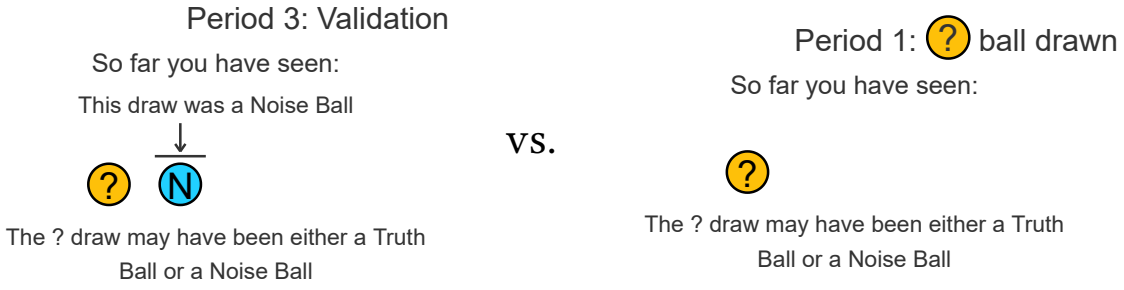
Friedman (1998) finds subjects err with striking consistency, often choosing to keep their choices.

[17] Related to this point, Guay et al. (2023) mentions that studies often obtain different results depending on whether they vary the extent to which subjects are shown exclusively fake news versus a mix.

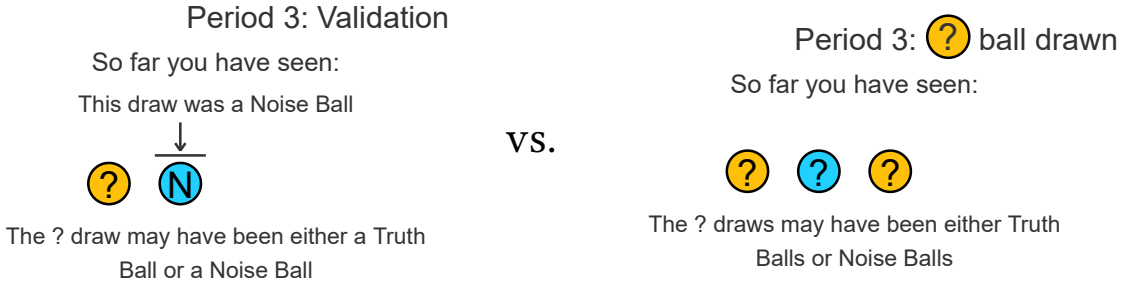
[18] In ongoing research we examine a version of this experiment using targeted (i.e., non-verifying) retractions; the results are largely consistent, although direct comparisons between the two are unwarranted. These results are available from the authors upon request.

(a) *Comparing beliefs with retractions and without the retracted observation*: Are subjects' beliefs after seeing a retraction the same as if the retracted observation had never been observed in the first place?

(b) *Comparing beliefs with retractions and with equivalent new observation*: Are subjects' beliefs following a retraction of a *yellow* signal the same as when observing a new *blue* draw?



(a) Retraction vs. No Retracted Draw



(b) Retraction vs. Equivalent New Draw

Figure 1: Illustrative examples to explain the empirical strategy

Notes: This figure provides an illustrative example of the empirical strategy for testing Hypothesis 1. According to Proposition 1, for any quasi-Bayesian, beliefs should be identical for each of the displayed histories. Panel (a), comparing beliefs following *(yellow, blue, retraction of blue)* to those following *(yellow)*, tests if beliefs are the same when evidence gets retracted and when such evidence was never observed. Panel (b), comparing beliefs following *(yellow, blue, retraction of blue)* to those following *(yellow, blue, yellow)*, tests whether updating from retractions is the same as from otherwise equivalent direct information.

While (a) and (b) can both be used to assess whether retractions are less effective, and although 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)—diminished updating from retractions could be driven by a feature of belief updating common to both retractions and new information. Conversely, new observations and retractions could be treated differently, but with retracted evidence treated as if it had never been seen, and with over-reaction to new observations driven by some other channel—leading to (b) without (a).

3.3. Estimation Strategy

We start by noting that belief updates in log-odds should be $\pm K(\textit{yellow})$, no matter the signal (a draw or a retraction) and no matter the prior (moderate or extreme).[19] This is because a Bayesian would have constant log-odds updates for any prior. Therefore, since using log-odds beliefs allows us to more easily compare and interpret our results, and in keeping with standard practice in the literature on belief updating (as in Benjamin, 2019), we will specify all our regressions using log-odds of beliefs, defined as $\hat{\ell}_t := \mathscr{L}(\hat{p}_t)$. Our conclusions are, however, robust to relying either on log-odds or level beliefs, as shown below.

The key element of our estimation strategy relies on precisely defining fixed effects based on the comparisons described in Proposition 1 (and illustrated in Figure 1) to identify diminished effectiveness of retractions. For this, we will pair histories $\mathscr{H}_t$ with and without retractions. Recall that $\mathscr{H}_t$ denotes the history up to and including period t: the set of all the draws observed and any retractions, fixing the order. Except for Section 4, where we explicitly consider updating *after* retractions, we do not include histories in which there was previously a retraction or where the truth ball was revealed so as to avoid any confounding factors.

For *comparing beliefs with retractions and without the retracted evidence*, test (a), we define the *compressed history*, $C(\mathscr{H}_t)$: the history with the retracted observations removed, as if they had never occurred to begin with. Taking as example the top panel of Figure 1, the compressed history of *(*yellow, blue, *retraction of the* blue*)* is simply *(*yellow*)*.[20] According to Hypothesis 1a—based on Proposition 1(a)—histories sharing a common compressed history should also share common beliefs and, therefore, the same log-odds beliefs.

We then test Hypothesis 1a with the following regression,

$$\hat{\ell}_{i,t} = \beta_0 \cdot r_{i,t} + \beta_1 \cdot r_{i,t} \cdot K(s_{i,\rho_{i,t}}) + \gamma_{C(\mathscr{H}_{i,t})} + \varepsilon_{i,t}. \tag{2}$$

where i denotes the subject, $r_{i,t}$ denotes a dummy variable indicating in period t there is a retraction ($r_{i,t} = 1$) or a new observation ($r_{i,t} = 0$), $s_{\rho_{i,t}}$ denotes the color of the retracted observation, $\gamma_{C(\mathscr{H}_{i,t})}$ are fixed effects for compressed history, and $\epsilon_{i,t}$ is a noise term.

The coefficient of interest is β_1. In the context of the illustrative example, our compressed-history fixed effects allow us to difference beliefs across histories that induce the same compressed history, *(yellow)*, such as *(yellow)* and *(yellow, blue, retraction of blue)*. As $K(\textit{blue}) < 0$, retracting *blue*

[19] In contrast, the change in levels is lower the farther away from 1/2 the prior belief is.

[20] 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*)*.

should increase the belief that $\theta = yellow$, and so β_1 captures how much less beliefs update from a retraction compared to how much they update from the retracted observation when it was first observed. Hypothesis 1a corresponds to $\beta_1 > 0$.[21]

For *comparing retractions to new evidence*, test (b), we define *sign history*, $S(\mathcal{H}_t)$, which is the history without distinguishing whether signals were new observations or retractions. For example, as illustrated in the bottom panel of Figure 1, *(yellow, blue, retraction of blue)* and *(yellow, blue, yellow)* both have the same sign history. We then run the same regression as before, Equation 2, except with sign-history fixed effects, $\gamma_{S(\mathcal{H}_t)}$, instead of compressed-history fixed effects, $\gamma_{S(\mathcal{H}_t)}$. β_1 again is the coefficient of interest, measuring how much less beliefs update from retractions than from (informationally) equivalent new observations.

3.4. Updating from New Observations

As a first step in our analysis, and in part as a test of the validity of our experimental setting, we examined subjects' belief updating from (non-retracted) new observations using a standard empirical approach in this literature. Here, we simply note that our findings are consistent with existing literature—we present the results more in-depth in Section 5, where we investigate how retractions affect belief updating patterns.

In the absence of a retraction, the design is similar to many others surveyed by Benjamin (2019). Subjects appear to correctly understand the setting, with reported beliefs tracking Bayesian posteriors closely.[22] 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. Specifically, we replicate common patterns in belief updating, such as base-rate neglect and confirmation bias. While subjects depart from Bayesian updating, our theoretical framework implies that any additional departure due to retractions cannot be attributed to explanations that are not specific to the nature of the information source. We first focus on how belief updating from retractions differs from updating from new observations, deferring the detailed reporting and discussion of general departures from Bayesian updating to Section 5.

[21] The scaling by $K(s_{\rho_t})$ will prove useful when discussing how much subjects infer from observations in the same log-likelihood scale to enable a comparison to Bayesian updating. Note that, upon observing s_t, Bayesian updating implies that $\ell_t = \ell_{t-1} + K(s_t)$, where $\ell_t := \mathscr{L}(p_t)$.

[22] Online Appendix B.1, presents beliefs and Bayesian posteriors disaggregated by history; in Online Appendix A, we report the difference and the distance between beliefs and Bayesian posteriors.

Retraction vs.	No Retracted Draw	Equivalent New Draw
	(1)	(2)
	$\hat{\ell}_t$	$\hat{\ell}_t$
Retraction (r_t)	0.011	-0.019
	(0.018)	(0.025)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.586^{***}	0.603^{***}
	(0.067)	(0.087)
Compressed History FEs	Yes	No
Sign History FEs	No	Yes
R^2	0.26	0.27
N	39162	39162

Clustered standard errors at the subject level in parentheses.

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

Table 2: Updating from Retractions (Hypothesis 1)

Notes: Column (1) tests Hypothesis 1a by estimating Equation 2. Column (2) tests Hypothesis 1b by estimating a variant of Equation 2, in which compressed-history fixed effects are replaced with sign-history fixed effects. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

3.5. Updating from Retractions

We now present our first central finding: empirical support of Hypothesis 1. We estimate the differences in beliefs specific to retractions using equation (2) on our baseline treatments.

We find a diminished effectiveness of retractions: subjects update beliefs less from retractions than from both the retracted observation (Retraction vs. No Retracted Draw) and an equivalent new observation (Retraction vs. Equivalent New Draw). Table 2 presents our estimates for our baseline treatments. Belief updates are significantly lower for retractions than new information: by 0.586 for retractions compared to belief updates had the retracted evidence never been observed and by 0.603 compared to equivalent new draws.

In order to contextualize this number, we compare it to the estimate of how much beliefs update following a new observation. This estimate is given by the coefficient β_1 from the regression specification $\Delta\hat{\ell}_{i,t} = \beta_0 + \beta_1 \cdot K(s_{i,t}) + \gamma_{S(\mathcal{H}_{i,t})} + \varepsilon_{i,t}$, where $\Delta\hat{\ell}_{i,t} = \hat{\ell}_{i,t} - \hat{\ell}_{i,t-1}$, restricted to histories $\mathcal{H}_{i,t}$ consisting only of new draws.[23] We find that a new draw moves beliefs by 1.081 times the log-likelihood of a new draw, providing a rough estimate of how much *less* beliefs update from retractions relative to new draws: .603/1.081, approximately 55%. Panel (a) of Figure 2 provides a visualization of these estimates. Panel (b) provides analogous estimates using levels ($\hat{p}_t$), instead

[23] Note that when we restrict to histories without retractions, compressed and sign histories are the same: $C(\mathcal{H}_t) = S(\mathcal{H}_t)$; hence, this normalization is appropriate for both comparisons.

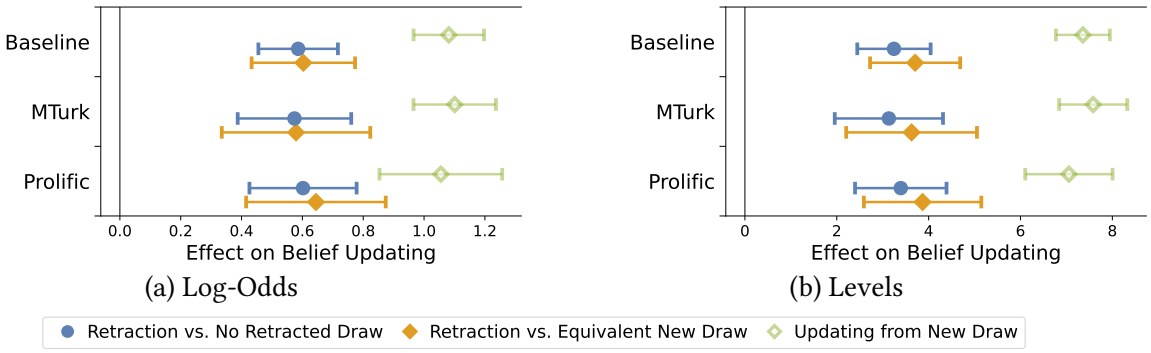


Figure 2: Retractions are Less Effective (Hypothesis 1)

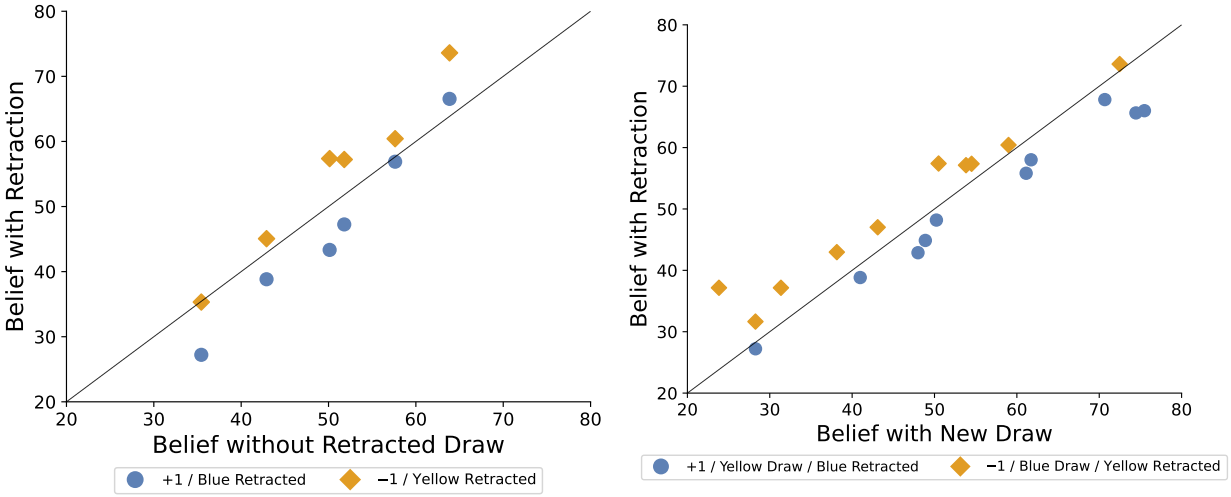
Notes: This figure depicts the effects of retractions on belief updating, showing how much less subjects update beliefs from retractions than from the retracted evidence (Retractions vs. No Retracted Draw; blue solid circle) and from new direct evidence (Retractions vs. Equivalent New Draw; orange solid diamond). The green hollow diamond depicts how much beliefs update on average from new draws, for comparison. Panel (a) shows these estimates for beliefs in log-odds ($\hat{\ell}_t$) as per Equation 2, while panel (b) provides the analogous estimates for beliefs in levels ($\hat{p}_t$). The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period. The figure displays results both pooled (Baseline) and separated by recruitment platform.

of log-odds, and exhibits consistent results. Specifically, we find that following retractions (i) beliefs update insufficiently and remain on average 3.2 percentage points away from the beliefs held absent the retracted evidence, and (ii) subjects update beliefs on average 3.7 percentage points less than from new draws—about 50% of the average belief updates from observations of 7.4 percentage points.

We conclude that subjects infer substantially less from retractions than direct evidence. Furthermore, this difference does not depend on whether test (a) or test (b) is considered.

These findings represent average estimates, and a natural question is the extent to which there is heterogeneity in the effects across histories. Throughout, we will discuss different meaningful dimensions of heterogeneity, namely with respect to how recent retracted observations are and the number of draws observed (Section 4.4), as well as if the retraction is confirmatory (reinforces the prior belief) or not (Section 5). While we lack statistical power at the most disaggregated level, Figure 3 provides indicative evidence that our results are robust across histories, and we also report results fully disaggregated by history, with consistent conclusions across histories (see Online Appendix B.2).

We note that we collected data for our baseline design twice, on Amazon Mechanical Turk in 2020 and again on Prolific in 2024. We obtained remarkably similar estimates of the effect across both platforms, as seen in Figure 2. In Appendix C we show there are no significant differences between the two recruitment platforms, across all our main specifications. We discuss the robustness of



(a) Retractions vs. No Retracted Draw

(b) Retractions vs. Equivalent New Draw

Figure 3: Retractions are Less Effective (Hypothesis 1)

Notes: This figure exhibits the effect of retractions on belief updating across the fixed effects used in our baseline specifications, reported in Table 2. Each marker in panel (a) represents average beliefs with a retraction (y-axis) and without the retracted draw (x-axis) for a specific compressed history. Analogously, each marker in panel (b) represents average beliefs with a retraction (y-axis) and with an equivalent new draw (x-axis) for a specific sign history. Blue dots correspond to cases in which a blue draw is retracted, and orange diamonds to those in which the retraction refers to a yellow draw. Retractions being less effective corresponds to blue dots being below the 45-degree line and orange diamonds above. The sample includes all observations of subjects in the baseline treatments, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

our results further in Section 6, only mentioning for now that restricting to particular rounds or to subjects that appear to perform better does not affect our conclusions.

4. Informational Complexity and Diminished Updating from Retractions

Having documented differences in beliefs in updating from retractions, we now turn to a discussion of mechanisms. We divide our analysis of possible mechanisms into two parts. In this section, we propose and analyze the hypothesis that retractions are less effective because they entail greater informational complexity. We defer to the following sections the discussion of alternative explanations which could plausibly generate our results—and show that they do not.

4.1. Retractions Provide More Complex Information

While the informational content (as captured by the log-likelihood) of a retraction is the same as that of a new observation, we argue that properties inherent to retractions render it more complex and lead to the observed diminished belief updates.

One such property refers to the kind of information retractions provide. In contrast to observations that provide direct evidence about the state (e.g., statements, trials, data), retractions provide only indirect information. To see this, note that retractions' meaning is obtained by informing about the quality or properties of direct evidence and are hence "one step removed" from the state relative to observations. Inference from retractions, therefore, necessitates an additional layer of contingent reasoning compared to observations, which renders them more complex. Indeed, there is abundant evidence that contingent reasoning renders problems more complex and explains deviations from optimality. These include failure to incorporate pivotality considerations in voting (Esponda and Vespa, 2021), neglecting correlation in information sources (Enke and Zimmermann, 2019), or in common value auctions (Eyster, Rabin and Vayanos, 2019). Even in very simple environments, an added layer of contingent reasoning entails a significantly greater propensity for suboptimal choices (Martínez-Marquina, Niederle and Vespa, 2019).

In our setup, this additional layer of contingent reasoning can be precisely seen using a simple causal model as given by a directed acyclical graph (Pearl, 2009). Figure 4 represents how θ, s_t, ϵ_t, and n_t are related, whereby an arrow from variable x to variable y means that x determines (in part) the value of y. We say that an observation s_t provides *direct information* about the state θ, since s_t is directly connected to θ, with θ directly influencing the distribution over the observation's realization. However, information obtained from a retraction—disclosing n_t—is only indirectly informative about θ, as θ and n_t are independent. Dependence emerges only through conditioning on s_t: information that an observation is or is not noise (n_t) is only informative about θ *contingent on* s_t. Pearl (2009) refers to such connections as *indirect*. Pearl and Mackenzie (2018) argue that this phenomenon—i.e., that independent variables can become correlated conditional on another variable—is responsible for several apparent logical paradoxes.[24]

In our subsequent analyses, we turn to measuring complexity and identifying its prominent association with updating from retractions.

[24]For instance, the Monty Hall problem is central among the paradoxes described by Pearl and Mackenzie (2018), connecting this observation to our discussion of Miller and Sanjurjo (2019) from Section 3.1. Other related phenomena are the observed difficulty people have in thinking through problems involving higher-order reasoning, expressed in aversion to compound lotteries (Abdellaoui, Klibanoff and Placido, 2015; Dean and Ortoleva, 2019) and in mistaken higher-order beliefs in strategic settings (Crawford, Costa-Gomes and Iriberri, 2013; Kneeland, 2015; Alaoui and Penta, 2016; Alaoui, Janezic and Penta, 2020).

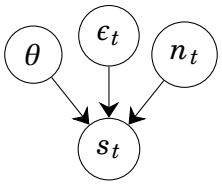


Figure 4: Graphical model representation of a new draw

4.2. Tracing Retraction Complexity

We now turn to our empirical measures of complexity. A common microfoundation for deviations from Bayesian updating is the hypothesis that the agent faces cognitive imprecision, as posited by models of cognitive uncertainty, efficient coding, and sequential sampling.[25] Our hypothesis is that this cognitive imprecision is higher for retractions. We provide evidence for this using two broad strategies. First, we consider different empirical measures of complexity borrowed from the literature and show that these generally are larger for retractions. Second, we consider treatments of and variation in our baseline design where retraction complexity would appear to increase, showing that this correspondingly strengthens the effect.

Before presenting our evidence for such a mechanism, we briefly sketch a model in the spirit of this literature, which ties complexity to empirical measures that we can infer from the data. Suppose decision-maker i faces uncertainty about how to interpret the likelihood of evidence $\boldsymbol{E}$ and update beliefs. In particular, for tractability, we assume the decision-maker's prior about θ is Gaussian, with $\boldsymbol{K}(\boldsymbol{E}) \sim \mathcal{N}(0, \sigma^2)$ and that they obtain T noisy estimates $\boldsymbol{K}(s_t) + \sigma_\zeta \cdot \zeta$, where ζ denotes Gaussian noise. This yields posterior log-odds updates as

$$\hat{\ell}_t = \hat{\ell}_{t-1} + \beta \boldsymbol{K}(\boldsymbol{E}) + \beta \frac{\sigma_\zeta}{\sqrt{T}} \zeta,$$

with $\beta = (1 + \sigma_\zeta^2/(\sigma^2 T))^{-1}$. Section 3 shows that β is lower for retractions. The hypothesis that retractions increase complexity is reflected in an increase of σ_ζ.

We test falsifiable predictions from this setup that could explain our results. For that, we use three behavioral markers of complexity: (1) *accuracy*, i.e., how close belief reports are to Bayesian posteriors; (2) *speed*, that is, decision times; and (3) *variability* in belief reports.

[25] 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, Li and Woodford (2021), Frydman and Jin (2022), Enke and Graeber (2022), and Augenblick, Lazarus and Thaler (2023). Models of sequential sampling provide a relationship between cognitive uncertainty and time through evidence accumulation (Krajbich, Armel and Rangel, 2010; Bhui and Gershman, 2018). See Ratcliff et al. (2016) for a survey of sequential sampling models in psychology and neuroscience, and Fudenberg, Strack and Strzalecki (2018), Alós-Ferrer, Fehr and Netzer (2021), and Gonçalves (2023) for recent applications in economics.

Accuracy. Our first indicator measures the distance between belief reports and the Bayes posterior. This variable captures accuracy since, based on our incentivization, the optimal report given the provided information coincides with the Bayesian posterior, and the expected payoff is decreasing in the belief bias, i.e., the distance between the belief reported and the Bayes posterior, $|\hat{p}_t - p_t|$.

Speed. Our second indicator captures how much effort individuals exert. A standard approach in the literature associates T with decision time, the idea being that the decision-maker obtains one such signal per unit of time spent deliberating (see footnote 25). In line with the general finding that decision-makers take more time and do less well on simple tasks when these tasks become less immediately apparent, we will interpret longer decision times, together with lower accuracy, as suggestive evidence that complexity is higher in the updating problem.[26]

Variability. Our third measure is the variability in the belief reports; following Khaw, Li and Woodford (2021) and Enke and Graeber (2021), we adopt it as an indicator of the underlying complexity. The underlying intuition is that greater cognitive imprecision generates less precise choices. In our model, given the above, an increase in σ_ζ^2 increases the variance of log-odds posterior beliefs insofar as the posterior variance about $K(E)$ is at most half of the prior variance about $K(E)$, i.e., σ^2—see Online Appendix C.1.

4.3. Retraction Complexity

The preceding discussion motivates the following hypothesis, which we proceed to analyze:

Hypothesis 2 (Retractions are More Complex)**.** *Inference from retractions is more difficult than processing new observations, resulting in (a) greater belief bias, (b) longer decision time, and (c) higher belief variance.*

To test this hypothesis, we use an identification strategy similar to the one used to test the effects of retractions on belief updating (Section 3.3). Specifically, in Table 3, we estimate versions of the following:

$$y_{i,t} = \beta_1 \cdot r_{i,t} + \gamma_{i,t} + \varepsilon_{i,t}, \tag{3}$$

where $y_{i,t}$ is a dependent variable and $\gamma_{i,t}$ are the relevant fixed effects, as in Section 3.3 and under the same sample restrictions.

Specifically, to test if belief bias is greater and decision times longer when subjects face a retraction, the dependent variable $y_{i,t}$ corresponds to subject i's belief bias ($|\hat{p}_{i,t} - p_{i,t}|$), and to log

[26] Early evidence for this observation can be found in, for instance, Banks, Fujii and Kayra-Stewart (1976), Buckley and Gillman (1974), or Ratcliff (1978); see Gonçalves (2024) for a formal treatment.

Retraction vs.	No Retracted Draw			Equivalent New Draw		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\lvert\hat{p}_t - p_t\rvert$	$\ln(T_t)$	$\text{Var}(\hat{\ell}_t \mid h_t)$	$\lvert\hat{p}_t - p_t\rvert$	$\ln(T_t)$	$\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	2.765^{***}	0.064^{***}	1.240^{***}	1.111^{***}	0.084^{***}	0.580^{***}
	(0.266)	(0.012)	(0.172)	(0.282)	(0.014)	(0.171)
Mean Decision Time		8.830			8.830	
Compressed History FEs	Yes	Yes	Yes	No	No	No
Sign History FEs	No	No	No	Yes	Yes	Yes
R^2	0.07	0.01	0.03	0.08	0.01	0.03
N	39162	39162	5236	39162	39162	5236

Clustered standard errors at the subject level in parentheses.

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

Table 3: Effect of Retractions on Complexity Indicators (Hypothesis 2)

Notes: This table provides estimates of the effect of retractions on three indicators of complexity, following Equation 3. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(3)) and (b) updating from a retraction vs. an equivalent new draw (Columns (4)-(6)). Columns (1) and (4) refer to the accuracy of belief updating, defined as the absolute difference between beliefs and Bayesian posteriors. Columns (2) and (5) refer to the speed of response, defined as log decision time. Columns (3) and (6) refer to the variability of updating, defined as subject-level history-contingent log-odds belief variance. Decision time is measured in seconds. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

decision time ($\ln(T_{i,t})$), respectively. We perform both comparisons outlined in Hypothesis 1: (a) retractions compared to histories where the draw was never observed, using compressed history fixed effects ($\gamma_{i,t} = \gamma_{C(\mathscr{H}_{i,t})}$), and (b) retractions versus an equivalent new draw, relying on sign history fixed effects ($\gamma_{i,t} = \gamma_{S(\mathscr{H}_{i,t})}$).

We test if retractions increase belief variance by taking the dependent variable to be the sample variance of beliefs computed at the subject level and conditional on (i) whether a retraction was observed and (ii) either the compressed history ($\text{Var}(\hat{\ell}_{i,t} \mid C(\mathscr{H}_{i,t}), r_{i,t})$), or the sign history ($\text{Var}(\hat{\ell}_{i,t} \mid S(\mathscr{H}_{i,t}), r_{i,t})$). Here, due to power considerations, we treat compressed/sign histories that are the same up to permutations as the same, and therefore, estimate within-subject belief variance at a given (permuted) compressed/sign history—for notational simplicity, we maintain the same notation.

Table 3 confirms Hypothesis 2. Retractions decrease accuracy in that belief bias increases both compared to not having seen the retracted draw (by almost 3 percentage points—Column (1)) and compared to an equivalent new observation (by over 1 point—Column (4)). Subjects also take longer in reporting beliefs—approximately 6% compared without the retracted observation

(Column (2)) and 10% longer when compared to an equivalent new draw (Column (5))—a conclusion which remains valid when controlling for experience and considering only later rounds.[27] Columns (3) and (6) provide an analogous comparison for the (log-odds) belief variance estimated at the subject level, which retractions increase significantly—by over one-third in either case. In both cases, we see that belief variance increases following a retraction. Figure 5 below provides a visualization of the results in Table 3.

Our results suggest that retractions are not only treated differently but also involve greater complexity. 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 a retraction's informativeness relative to direct information about the state of the world.

4.4. Validating and Varying Complexity

We now show that variation in the strength of belief updating moves together with predictions that would emerge from a complexity-based mechanism. In particular, we complement our analysis by assessing whether, in situations that we would expect to be more complex, our proxies for complexity are aligned, and if beliefs are correspondingly less responsive to more complex information.

We first exploit the natural variation in our experimental design to consider cases in which retractions should be less complex. If, at time t, the observation received at $t-1$ is retracted, subjects need only to revert to the belief they held at $t-2$, that is, before receiving that observation. In contrast, inferring from a retraction of previous evidence involves forming beliefs about a dataset not previously observed, thus involving counterfactual reasoning. Hence, we expect retractions of more recent observations to be easier to process than retractions of less recent observations, and, consequently, more effective in moving beliefs:

Hypothesis 3 (Retracting Recent Observations is Easier)**.** *Retractions of recent observations are (a) more effective and (b) less complex.*

To assess Hypothesis 3, we use the same regression specifications and contrast the estimates of the effect of retractions on belief updating (Table 2), belief bias, decision time, and belief variance (Table 3) in our baseline treatments to the estimates one obtains when considering only retractions of the more recent observation.

[27] See Section 6.2. While our results show subjects take less time in later rounds, the increase in decision time caused by retractions remains consistent in later rounds, when subjects have had more experience observing retractions. 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, it appears unsurprising that we do not detect a difference depending on whether subjects has seen more retractions in the past.

We also examine how retractions affect inference from subsequent new evidence. Our posited mechanism suggests that if a retraction is harder to process, then it may be more difficult to update following a retraction. To see why, we note that a signal history S_t will generally influence how a subject should respond to s_{t+1} via its implications on θ; the added complexity of retractions would then imply spillovers as subjects would correspondingly face greater difficulty understanding what this implication should be. This idea underlies another expression of our proposed mechanism, which we articulate as a related hypothesis:

Hypothesis 4 (Updating after Retractions)**.** *Following a retraction, (a) subjects update less from new observations, and (b) inference is more difficult.*

Since subjects update differently from a retraction than from an equivalent new draw, a difference in beliefs $\hat{\ell}_t$ following a retraction in period $t-1$ may just be an expression of the difference in the history at $t-1$. In order to test if subjects update less after a retraction, one needs to now explicitly consider how the *change* in log-odds beliefs at a particular sign history is affected by having observed a retraction in the previous period. We do this by estimating the following: $\Delta\hat{\ell}_{i,t} = \beta_0 + \beta_1 \cdot r_{t-1} + \gamma_{S(\mathcal{H}_{i,t})} + \varepsilon_{i,t}$, where $\gamma_{S(\mathcal{H}_{i,t})}$ denotes sign-history fixed effects, and $\Delta\hat{\ell}_{i,t} := \hat{\ell}_{i,t} - \hat{\ell}_{i,t-1}$ the change in log-odds beliefs. To test Hypothesis 4b, we consider an analogous version of equation (3): $y_{i,t} = \beta_1 \cdot r_{i,t-1} + \gamma_{S(\mathcal{H}_{i,t})} + \varepsilon_{i,t}$, where $y_{i,t}$ is a dependent variable. We exclude periods in which the truth ball was revealed for obvious reasons.

We find support for both Hypotheses 3 and 4. As shown in Figure 5(a), retractions of more recent observations are significantly more effective. Specifically, subjects updates about 35-40% less from retractions of recent draws than from equivalent new draws, in contrast to approximately 50-55% in our baseline. In line with greater effectiveness, we find that belief reporting is starkly faster when the retraction refers not to an earlier but to the last draw (panel (c)), and also that retractions of more recent observations induce lower belief variances (panel (d)). We further observe that belief bias is attenuated (panel (b)), although not significantly different from our baseline in one case. Regarding Hypothesis 4, we find that subjects update less after retractions than after equivalent new draws (a), are more biased (b), take longer (c), and exhibit higher variability in their reports (d).

To summarize, consistent with our posited mechanism, the data suggests retractions of more recent observations are less cognitively demanding and that inference from new draws is more complex if they follow a retraction. In both cases, the intensity of belief updates aligns with our complexity indicators.

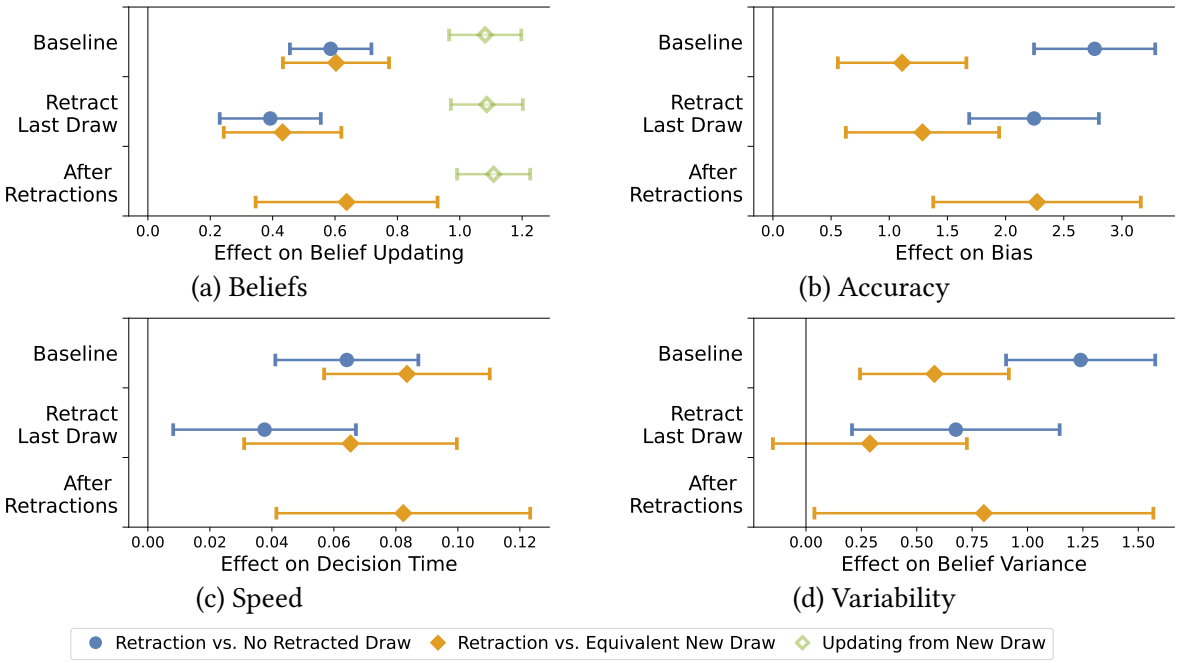


Figure 5: Retracting Recent Evidence and Evidence After Retractions (Hypotheses 3 and 4)

Notes: This figure provides estimates for the effect of retractions on belief updating and on three complexity indicators, across settings in which we expect complexity to change. "Retract Last Draw" restricts the sample of retractions to retractions in which the most recent draw is retracted, corresponding to Hypothesis 3. "After Retractions" considers updating from new draws contingent on whether or not a retraction occurred in the past, corresponding to Hypothesis 4. Panel (a) displays the effect of retractions on belief updating, $\hat{\ell}_t$, under the same specifications as for Figure 2. Panels (b)-(d) display effects on our three complexity indicators—accuracy ($|\hat{p}_t - p_t|$), speed ($\ln(\mathrm{T}_t)$), and variability ($\mathrm{Var}(\hat{\ell}_t \mid h_t)$)—under the same specifications as for Table 3.

5. Belief Updating Patterns under Retractions

So far, we have provided evidence that complexity considerations can explain the diminished effectiveness of retractions. Here, we discuss how retractions entail signficantly different belief updating patterns compared to updating from new direct evidence.

While our results imply that retractions—indirect information—are treated differently from direct information, one possibility is that retractions simply magnify known updating biases. To examine this, we rely on Grether (1980) log-odds regressions, the main workhorse in the existing literature (cf. Benjamin, 2019), estimating variants of the following:

$$\hat{\ell}_{i,t} = \beta_0 + \beta_1 \cdot \hat{\ell}_{i,t-1} + \beta_2 \cdot K_{i,t} + \beta_3 \cdot K_{i,t} \cdot c_{i,t} + \varepsilon_{i,t}, \tag{4}$$

where $\hat{\ell}_{i,t}$ denotes i's log-odds belief at period t, $K_{i,t}$ the log-likelihood of the signal—i.e., $K(s_{i,t})$ in the case of a new draw $s_{i,t}$, and $-K(s_{i,\rho_{i,t}})$ for retractions—and $c_{i,t}$ an indicator variable that equals 1 whenever the signal observed confirms the prior belief ($\mathrm{sign}(\ell_{i,t-1}) = \mathrm{sign}(K_{i,t})$) and

	(1) $\hat{\ell}_t$	(2) $\hat{\ell}_t$
Signal (K_t)	1.102***	0.907***
	(0.060)	(0.060)
Prior (l_{t-1})	0.801***	0.747***
	(0.032)	(0.032)
Confirmatory Signal ($K_t \cdot c_t$)	–	0.651***
		(0.097)
Retraction (r_t) x Signal (K_t)	-0.768***	-0.516***
	(0.071)	(0.074)
Retraction (r_t) x Prior (l_{t-1})	0.042	0.106***
	(0.037)	(0.039)
Retraction (r_t) x Confirmatory Signal ($K_t \cdot c_t$)	–	-0.807***
		(0.130)
R^2	0.42	0.42
N	39162	39162

Clustered standard errors at the subject level in parentheses.

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

Table 4: Belief Updating Patterns under Retractions: Grether Regressions

Notes: This table shows that patterns in belief updating from retractions do not simply reflect a strengthening of known updating biases. It reports estimates of Equation 5 interacting the independent variables with whether or not the signal was a retraction (r_t). The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

0 if otherwise. Bayesian updating implies that $\beta_1 = 1$, $\beta_2 = 1$, and $\beta_3 = 0$. Base rate neglect, for instance, corresponds to $\beta_1 < 1$; under- and over-inference are expressed by $\beta_2 < 1$ and > 1, respectively; and confirmation bias, to updating relatively more from signals when these confirm one's prior belief, i.e. $\beta_3 > 0$.

In examining how patterns in updating from retractions differ from updating from direct evidence, we fully interact the specification given above with the dummy variable r_t indicating whether or not the signal corresponds to a retraction or a new draw:

$$\hat{\ell}_{i,t} = \beta_0 + \beta_1 \cdot \hat{\ell}_{i,t-1} + \beta_2 \cdot K_{i,t} + \beta_3 \cdot K_{i,t} \cdot c_{i,t} + r_{i,t} \cdot [\gamma_0 + \gamma_1 \cdot \hat{\ell}_{i,t-1} + \gamma_2 \cdot K_{i,t} + \gamma_3 \cdot K_{i,t} \cdot c_{i,t}] + \varepsilon_{i,t}. \quad (5)$$

The interaction terms allows us to examine how previously documented deviations from Bayesian updating vary depending on whether or not the signal is a retraction. Table 4 presents these results.

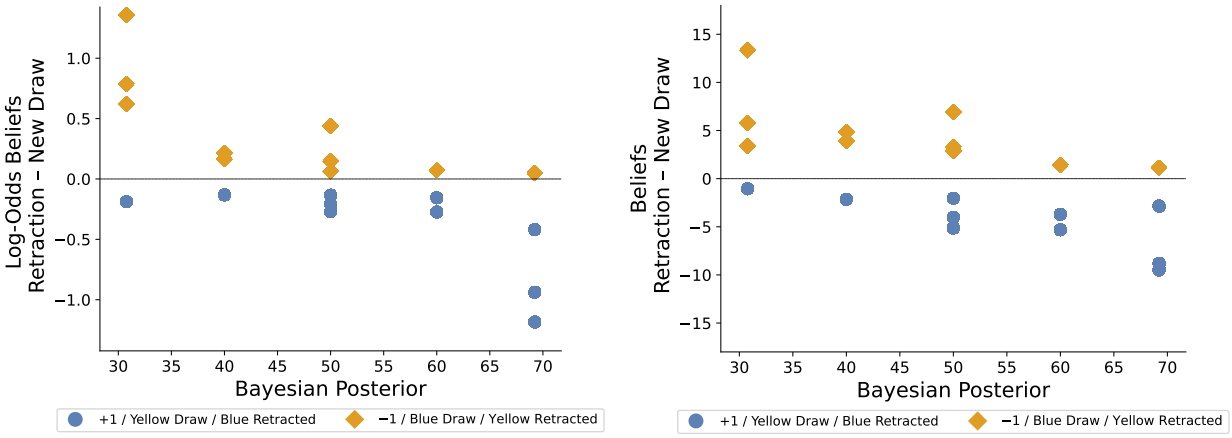
As foreshadowed in Section 3.4, we replicate known updating patterns. In line with results by

Augenblick, Lazarus and Thaler (2023),[28] we find $\hat{\beta}_2 = 1.102$, indicating weak over-inference from new observations, although not statistically different from 1. Once we consider whether the signal is confirmatory, we then obtain under-inference from new observations, with $\hat{\beta}_2 = 0.907$ and not statistically different from 1, while $\hat{\beta}_3 = 0.651 > 0$ indicates confirmation bias, resulting in over-inference from confirmatory information ($\beta_2 + \beta_3 > 1$)—a phenomenon previously documented by, e.g., Charness and Dave (2017). Together, this finding suggests that our subjects slightly overreact to new observations. However, this conclusion is primarily driven by confirmation bias: subjects update more from a signal when it corroborates their prior belief. 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 $\beta_1 < 1$.

A striking difference emerges: while updating from new draws exhibits slight over-inference ($\beta_2 \geq 1$) driven by confirmation bias ($\beta_3 > 0$), updating from retractions leads to marked *under*-inference ($0 < \beta_2 + \gamma_2 < 1$) and *anti*-confirmation bias ($\beta_3 + \gamma_3 < 0$). In sum, belief updating from retractions exhibits biases opposite those that emerge when updating from new draws, a conclusion which is robust across specifications. This nuance strengthens our finding that retractions are treated differently from new signals, as the behavioral responses to retractions are not simply accentuating pre-existing biases. In fact, retractions induce opposite biases in belief-reporting behavior.

These results suggest a specific form of heterogeneity in the diminished effect of retractions across different histories. We examine this heterogeneity using our baseline identification strategy (Section 3.3). Figure 6 shows the difference between beliefs updated from retractions and equivalent new draws for each sign history. In line with the documented expression of anti-confirmation bias in Table 4, subjects update less from confirmatory retractions than from confirming new draws at extreme histories, following which they hold more extreme beliefs. Table 4 documents (1) a general diminished updating from retractions relative to new draws, (2) confirmatory bias from new draws, and (3) anti-confirmatory bias from retractions. Figure 6 illustrates this finding: it is exactly at more extreme histories, entailing more extreme beliefs, when observing a confirmatory signal induces subjects to update less from retractions relative to new draws, as (2) and (3) there enhance (1). In contrast, (2) and (3) counter (1) for disconfirmatory signals, explaining why the difference between beliefs following retractions and new draws is small in this case, even if with the anticipated sign. We emphasize that this result does not speak to which beliefs are more

[28] Augenblick, Lazarus and Thaler (2023) 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.



(a) Log-Odds (b) Levels

Figure 6: Updating from Retractions: Heterogeneity by History

Notes: This figure displays the difference between beliefs following retractions versus equivalent new draws disaggregated by sign history. Blue circles represent sign histories in which the last signal was either the retraction of a blue draw or a new yellow draw. Orange diamonds represent sign histories in which the last signal was either the retraction of a yellow draw or a new blue draw. Panel (a) presents results in log-odds, while panel (b) presents results in levels. In both cases, the x-axis is the Bayesian posterior of the sign history. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

difficult to update from[29]—rather, it speaks to the differential impact of retractions.

6. Robustness of the Findings

We performed extensive robustness checks to assess the validity of our results. In this section, we examine the extent to which our results (1) are driven by subject understanding, (2) reflect a general feature of behavior or rather depend on specific individual characteristics, and (3) are affected by design choices.

6.1. Robustness 1: Subject Screening and Understanding

Subject Screening. We strove to ensure that our results were not driven by inattentive subjects. While the behavior of participants on Amazon Mechanical Turk and Prolific has been shown to approximate well representative-population samples, it can sometimes be 'noisy' relative to traditional laboratory subjects (Snowberg and Yariv, 2021; Gupta, Rigotti and Wilson, 2021). To ensure

[29]Indeed, evidence for our complexity indicators is mixed, suggesting one should not infer that the heterogeneity across histories is motivated by varying degree complexity in updating. While we do find that decision time patterns by sign history are strongly related to those in Figure 6, the difference in bias when updating from retractions and new draws, however, is greater for histories inducing more moderate posteriors, and a similar phenomenon seems to occur with belief variability—see Online Appendix D.2.

our data was of high quality, we restricted participation to US residents with high approval rates (over 95%) and held our study during business hours (Eastern Standard Time), added captchas throughout the experiment, employed an incentivization scheme involving a high baseline and reward pay (see Section 2.3), and precluded the possibility of repeating the experiment. Additionally, we included comprehension questions in the instructions, which subjects had to answer correctly to proceed. These quality checks were important for us to be able to meaningfully test our hypotheses. If subjects were simply answering randomly, they would be biased relative to Bayesian updating but would exhibit no difference between updating from retractions relative to direct evidence.

Subject Understanding. We further examined the robustness of our results to excluding subjects based on different measures of inattentiveness. The results are robust, and if anything slightly stronger, when restricting the sample to those subjects who appear attentive, as defined in four different ways. First, using the comprehension questionnaire, we restrict our sample to subjects who answered all questions correctly on their first try ("Comprehension Correct"). While unincentivized, the majority of the subjects demonstrated clear understanding: approximately 60% and 90% answered all questions correctly on the first and second try, respectively; when answering randomly, the probability of answering all correctly on the first try would be 0.2% (see Appendix D). Second, we further restrict the sample to subjects who, when the state is revealed, correctly report that they know the state ("Understands Disclosure"). 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 ("Fewer Mistakes").[30] Fourth, we exclude subjects who could be mistaking sampling with and without replacement ("Understands Replacement").[31]

In Figure 7, we exhibit the estimates of the coefficient of interest corresponding to our baseline tables (2 and 3); the supporting regression tables can be found in Online Appendix E.1. The robustness of the results is consistent with noisy subjects if anything attenuating the effect, and shows that inattention is not driving our results.

Subject Confidence. We examine the possibility that retractions are associated with lower confidence, which would express greater cognitive uncertainty. For this, we included a question regarding subject confidence in all treatments in experiment C: similar to Enke and Graeber (2022),

[30] We considered various degrees of mistake-propensity: 1%, 5%, 10%, 20%; our conclusions remain the same. We also note that these checks are correlated. For example, the first two samples contain a substantially smaller fraction of subjects with excessively noisy reports.

[31] 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.

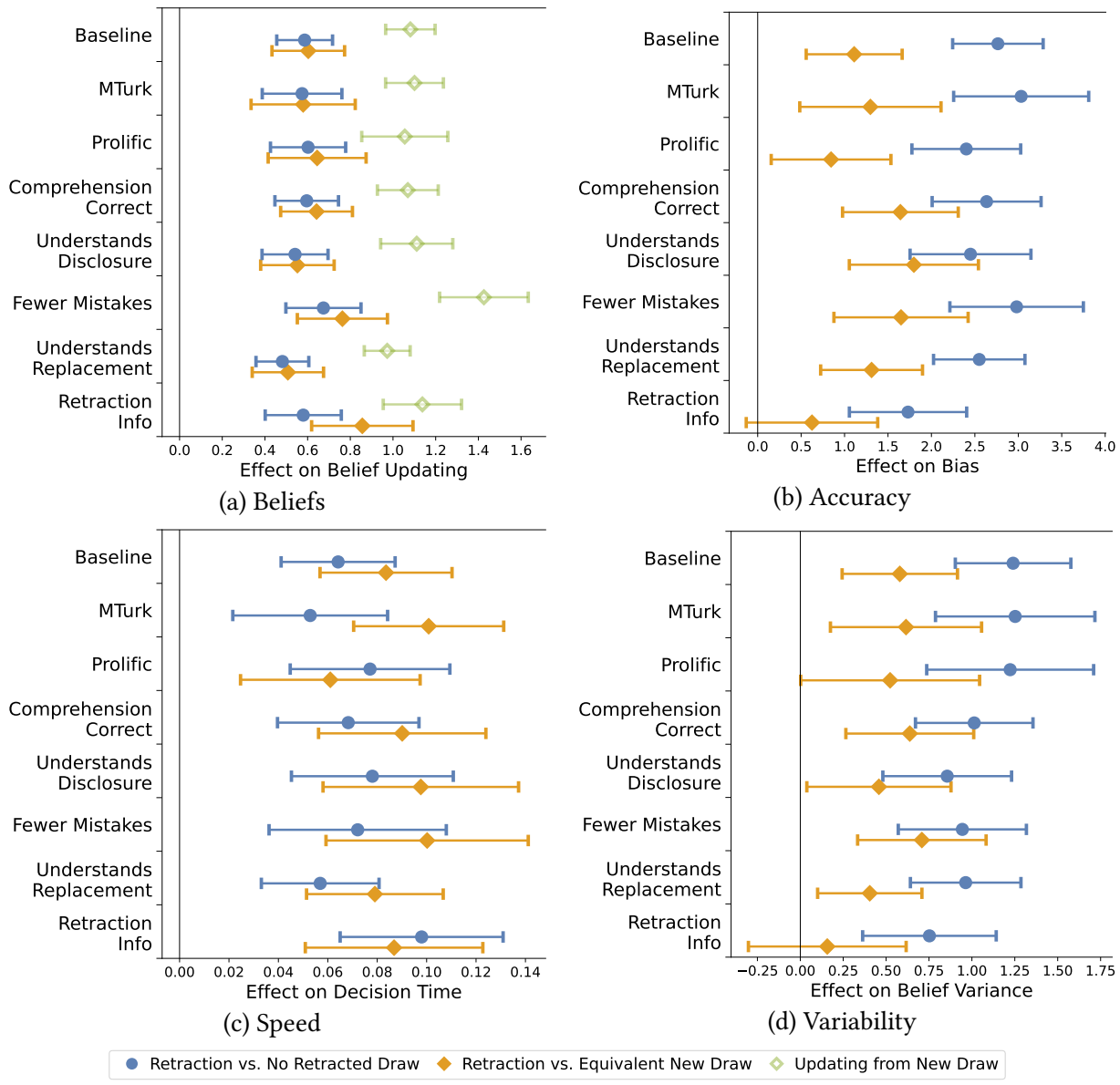


Figure 7: Robustness 1: Subject Screening and Understanding

Notes: This figure provides estimates of the effect of retractions on belief updating and on our three complexity measures across sample restrictions and experimental variants designed to test robustness to subject understanding. "Baseline" is the pooled sample of our baseline treatments; "MTurk" and "Prolific" split the sample by those platforms. We restrict the baseline sample to subjects who appear attentive in four ways: "Comprehension Correct" restricts to subjects who answered all experimental comprehension questions correctly on their first try; "Understand Disclosure" restricts to subjects who, when the state is revealed, correctly report that they know the state; "Fewer mistakes" removes subjects who update in the opposite direction to the signal more than 10% of the time; "Understands Replacement" excludes subjects who could be mistaking sampling with and without replacement. "Retraction Info" reports results from experiment C, where subjects are told retracted observations should be ignored. Panel (a) displays the effect of retractions on belief updating, $\hat{\ell}_t$, under the same specification as for Figure 2. Panels (b)-(d) display effects on our three complexity indicators—accuracy ($|\hat{p}_t - p_t|$), ($\ln(T_t)$), and variability ($\text{Var}(\hat{\ell}_t \mid h_t)$)—under the same specifications as for Table 3.

following the input of a belief report of $\hat{p} \in [0, 100]$, we ask subjects "Out of 100, how certain are you that the optimal estimate of the Truth Ball being yellow lies between $\hat{p}-1$ and $\hat{p}+1$?" Subjects then report a value between 0, labelled 'completely uncertain,' and 100, 'completely certain.'

In line with Enke and Graeber (2022), higher confidence is associated with subjects inferring more from new draws. However, this effect seems to be driven by greater confidence being associated with greater reliance on *confirmatory* signals.[32] Furthermore, confidence increases from approximately 60 out of 100, on average, to about 90 out of 100 when the truth ball is disclosed—a figure that is even closer to 100 for any of the sample restrictions discussed above.

While we do not find a significant difference in updating from retractions and direct evidence (see Table 27) depending on whether subjects are more or less confident, we do observe an effect of updating from retractions (relative to new draws) on subject confidence, albeit a small one: about 2 'confidence points' on average, and about 0.1 standard deviations in confidence, normalized within-subject (see Online Appendix E.2). This suggests that subjects are aware of, but ultimately underestimate, the greater complexity associated with updating from retractions, indicating—in the terminology of Enke, Graeber and Oprea (2023*b*) and Enke and Shubatt (2023)—that objective complexity (as revealed by behavior) is more severe than subjects' subjective perception, as given by reported confidence or cognitive certainty. Still, this finding provides reassurances that our results are not driven by subjects being uncomfortable with retractions or considering their interpretation insufficiently clear.

Additional Retraction Information. We examine if providing additional information about retractions improves outcomes significantly. In experiment C, we included a treatment ("Retraction Info") in which, when presented with a retraction, subjects are not only informed that a particular earlier draw was a noise ball but also told that "A noise ball is not informative about the color of the Truth Ball and you should ignore that you have seen it." The treatment is identical to our baseline but for this extra information. While some outcomes seem to improve (e.g. bias decreases, as does belief variance, and confidence in updating from retractions increases), the differences with respect to our baseline are not statistically significant—see Figure 7 and Online Appendix E.3. This suggests subjects err in interpreting the indirect evidence provided by retractions.[33]

[32] Specifically, we find that subjects that are, on average, more confident than the median infer slightly more, especially from confirmatory new draws. Perhaps more interesting is that when *within*-subject confidence is higher—i.e., using measures of confidence normalized for each subject—subjects do infer significantly more from signals, even though, again, this effect seems to be driven by inference from confirmatory signals. However, we find no significant correlation between confidence and bias—if anything, there is a weak positive correlation.

[33] This pattern is also reminiscent of findings documented in experimental tests of the Monty Hall problem, where individuals often fail to recognize the error even when told the correct way to reason through it (e.g. Friedman,

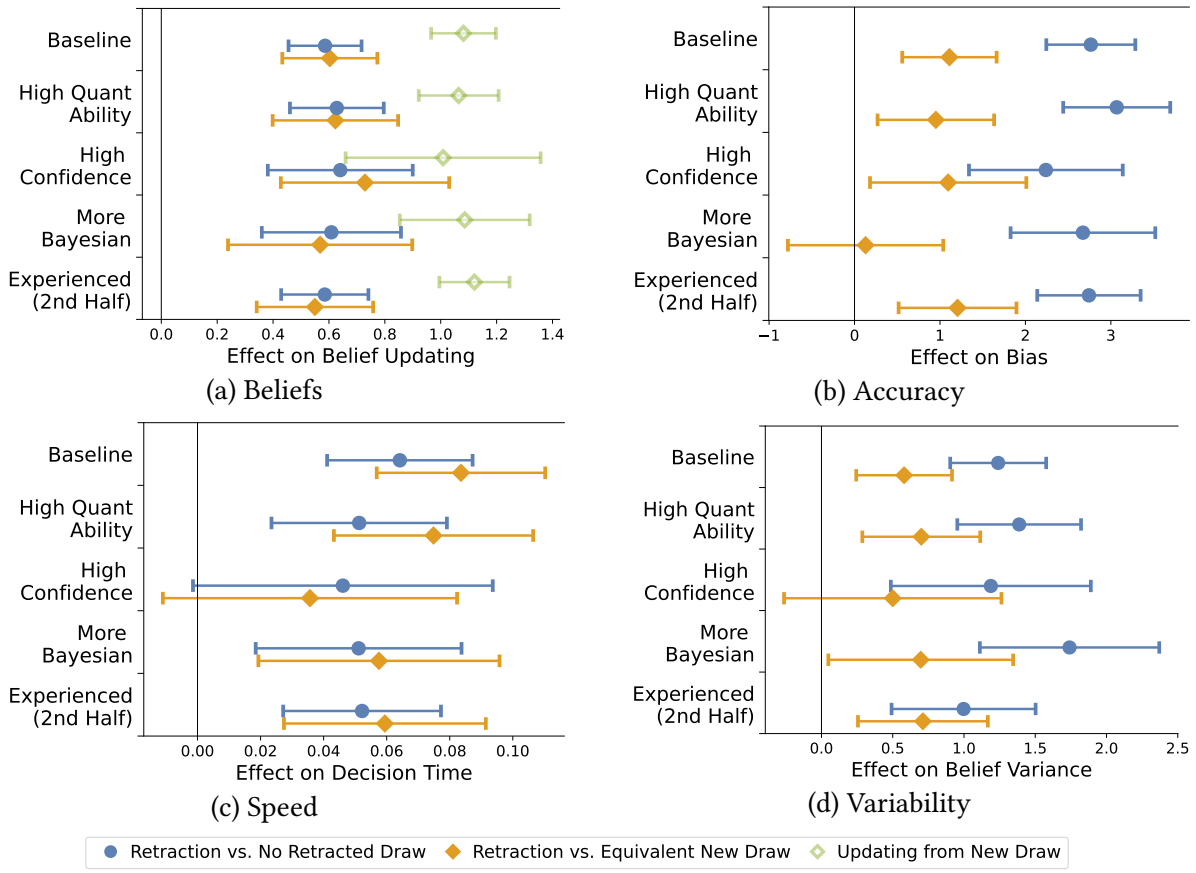


Figure 8: Robustness 2: Consistency Across Heterogeneity

Notes: This figure provides estimates of the effect of retractions on belief updating and on our three complexity measures across sample restrictions designed to test for heterogeneity. "Baseline" is the pooled sample of our baseline treatment. We restrict the baseline sample to test for heterogeneity in four ways: "High Quant Ability" restricts to subjects with above median score on a quantitative test in the experiment; "High Confidence" restricts to subjects with above median confidence in their beliefs; "More Bayesian" restricts to those who are more Bayesian than the median subject when updating from new draws; "Experienced" restricts to the second half of rounds for each subject. Panel (a) displays the effect of retractions on belief updating, $\hat{\ell}_t$, under the same specification as for Figure 2. Panels (b)-(d) display effects on our three complexity indicators—accuracy ($|\hat{p}_t - p_t|$), speed ($\ln(T_t)$), and variability ($\text{Var}(\hat{\ell}_t \mid h_t)$)—under the same specifications as for Table 3.

6.2. Robustness 2: Consistency Across Heterogeneity

Heterogeneous Treatment Effects. We explore heterogeneity in updating from retractions across multiple dimensions. We consider heterogeneity by whether subjects (i) have higher quantitative ability, as proxied for by their scores on incentivized quantitative multiple-choice questions which were asked at the end of the experiment ("High Quant Ability"); (ii) are more confident on average than the median subject ("High Confidence"); and (iii) are on average closer to the Bayesian posterior when updating from new draws than the median subject ("More Bayesian"). We re-estimate our main specifications on these groups (Figure 8) and expand our main specifi-

cations with interaction terms to account for heterogeneity (Online Appendix F.1), failing to find any relevant deviations from our baseline.

We also examine whether experience with the task affects our results. For this, we perform a similar heterogeneity analysis considering the second half of the experiment (rounds 17-32), at which point almost all subjects will have encountered a retraction. Again, we find no significant difference.

Finally, attesting to the robustness of our findings, we highlight that we replicated results using our baseline treatment in two different recruitment platforms, Amazon Mechanical Turk and Prolific, two years apart (see Appendix C).

Individual Heterogeneity. Underinference from retractions appears to be a robust feature within our sample, reflecting the overwhelming majority of subjects' behavior rather than a small minority. To show this, we estimate the specifications in Table 2 at the *subject* level. We report summary statistics on the subject-level estimates of the coefficient of interest (Retracted draw) in Online Appendix F.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.

That said, the following observations are notable: First, the estimates are strictly negative for most subjects (approx. 70%). Second, the mean estimate is higher than the median; thus, while most subjects infer less from retractions (with the median subject's bias still substantial), the distribution is skewed. Bootstrapped standard errors for both mean and median coefficients of interest show that these estimates are several standard deviations above 0, implying that these estimates are sufficiently precise to conclude that the diminished effectiveness of retractions is the rule, not the exception, among our subject pool. Finally, the individual-level estimates are single-peaked around the mean, pointing to a continuous spectrum of intensity of diminished inference from retractions rather than clearly distinguishable heterogeneous types.

6.3. Robustness 3: Variations on the Design

We now discuss our experimental design. We begin by revisiting how it contributes to our identification of mechanisms and subsequently examining the robustness of our results with respect to various design features.

1998). Pearl and Mackenzie (2018) discuss famous anecdotal instances of sophisticated individuals unwilling to admit errors in paradoxes involving reasoning with colliders.

6.3.1. Alternative Explanations Ruled Out by Design

We first take stock of alternative explanations for retraction failure that we rule out based on the design itself.

First, our use of a balls-and-urns design was motivated by our desire to tie the limited effectiveness of retractions to belief updating itself, minimizing the role of explanations related to particular domains (e.g., scientific understanding or political preferences). The fact that motivated reasoning is often at play in political domains might suggest it plays a crucial role in the limited effectiveness of retractions. While it could magnify it, we find this effect even without motivated reasoning. Additionally, even if we recognize memory is bound to play an important role in many settings, our baseline design also precludes memory-based explanations for retraction's limited effectiveness, as all information remained on the screen making the recollection of past signals simple.[34] Furthermore, issues of whether retractions lead to questioning the source's reliability, while interesting in their own right, are also precluded in our setting: a Bayesian decision-maker should be able to update beliefs from retractions without any ambiguity.[35]

Second, as Proposition 1 demonstrates, only explanations specific to retractions can rationalize retraction's diminished effectiveness. Indeed, we designed the experiment to compare retractions to informationally equivalent direct evidence. The paradigm we build on allows us to quantify objectively correct beliefs, which is difficult or impossible in domains where beliefs are subjective or, perhaps more problematically, not concretely defined. We can thus distinguish retraction failures from any explanation that applies to all forms of information processing and belief updating, such as confirmation bias. 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 5: biases in updating from retractions are not simply accentuated versions of known biases.

6.3.2. Variations on the Design

We ran several variations of our baseline design as different treatments in our four experiments. We discuss each of them, referring to our summary Figure 9 and additional analysis in Online Appendix G.

Elicit at the End. An alternative explanation for the diminished effect of retractions is that it is difficult to disregard evidence that has been actively used, as might be suggested by explanations

[34] Ratcliff's (1978) seminal paper already provided evidence that recall is imperfect even when referring to very short periods of time—and the more so, the greater the elapsed time.

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

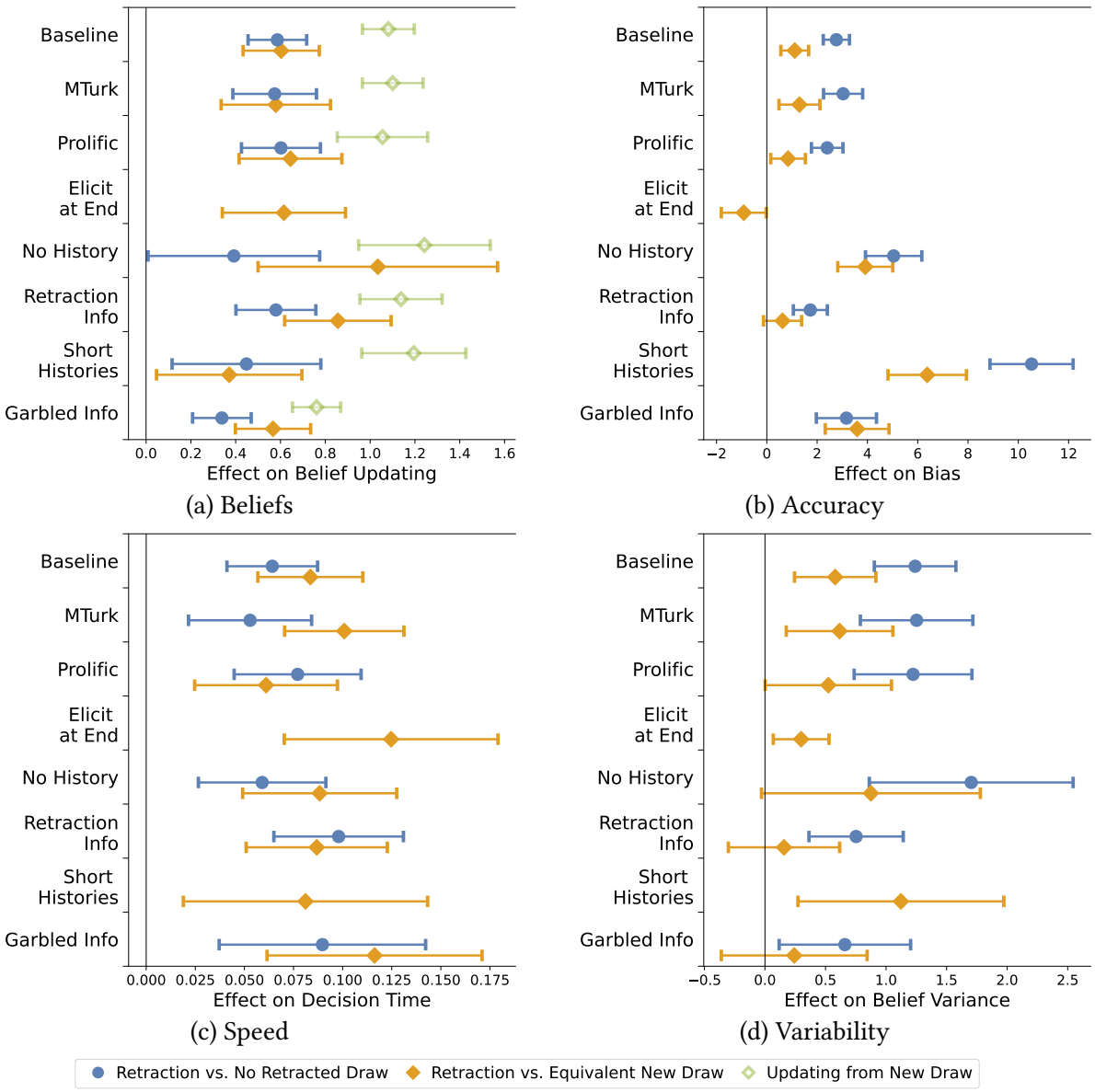


Figure 9: Robustness 3: Variations on the Design

Notes: This figure provides estimates of the effect of retractions on belief updating and on our three complexity measures across multiple variations on the baseline design. "Baseline" is the pooled sample of our baseline treatment; "MTurk" and "Prolific" split the sample by those platforms. In the "Elicit at End" variant, beliefs are elicited only at the end of each round. In "No History", subjects were only shown the current observation, not the history of all observations in the current round. In "Short Histories", there were only two periods per round, rather than four. In "Garbled Information", truth balls were not fully informative. Panel (a) displays the effect of retractions on belief updating, $\hat{\ell}_t$, under the same specification as for Figure 2. "Retraction Info" is described in Figure 7 and included to facilitate comparison to other treatments. Panels (b)-(d) display effects on our three complexity indicators—accuracy ($|\hat{p}_t - p_t|$), speed ($\ln(\mathrm{T}_t)$), and variability ($\mathrm{Var}(\hat{\ell}_t \mid h_t)$)—under the same specifications as for Table 3.

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 beliefs across our baseline—in which beliefs are elicited every period within a round—the "Elicit at End" treatment in experiment A—in which beliefs are elicited only at the end of each round.[36] The difference is null: having acted upon a piece of information or not does affect how much less one updates from retractions relative to equivalent new draws. Interestingly, bias in updating is lower, but a heterogeneity analysis reveals it to be only marginally significant (Table 34). While this does not imply that retractions are as (in)effective when individuals act upon past information in other contexts, it does strengthen our conviction that our results are not due to design details.

No History of Past Draws. It is often the case that, in real-world settings, past evidence remains available even if invalid, and retractions (e.g., of academic papers by journals or of news reports by media outlets) do not simply remove incorrect information but also describe what was corrected. Nevertheless, in many cases, the full history of past evidence may not be readily available either; it will necessarily be less salient and require being recalled. It is, therefore, natural to ask how omitting the history of past draws affects our baseline results. To speak to this, in our treatment "No History", the interface was kept exactly the same as in our baseline, except that the screen only showed the ball that had just been drawn and no other draws. When presenting retractions, we showed the retracted ball with the noise label, as in the original design, without any other draws. It was unclear if this would prompt subjects to misinterpret retractions as evidence for the opposite state and therefore lead to treating retractions as *more* informative than new draws and thus to updating more, not less, from retractions.

While removing the history does not result in statistically significant differences from our baseline in terms of how subjects update from retractions relative to new draws (Table 36), the data suggest that retractions become harder to interpret and that subjects update even less from retractions relative to new draws. Interestingly, removing the history of draws leads to notably higher variability in beliefs and more bias, resulting in less precise estimates for retraction effectiveness. Note that we would not expect to find this effect if subjects only paid attention to the last draw observed—the only piece of evidence necessary to update beliefs—suggesting that theoretically redundant past evidence plays a role in belief formation.

[36] Specifically, the "Elicit at End" treatment consisted of a sequence of events identical to 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 can nevertheless observe beliefs after two draws, as well as in period 3, whether there is a third draw or a retraction.

Short Histories. In a follow-up experiment, D, we presented subjects with an updating task identical to our baseline, except that in this treatment—labelled "Short Histories"—histories were shorter and ran for two periods only: subjects were provided one new draw in the first period, with the second signal being either a retraction or new draw. Our findings are robust even for short histories: subjects infer less from retractions, take longer, and are more biased when updating from retractions than from equivalent new draws, and the variability of beliefs is also higher. Although direct comparisons to our baseline are not well-founded, as these would partly reflect the documented heterogeneity of effects across histories (Section 5), we feel compelled to comment on the similarities and differences. The effect of retractions on diminished belief updating and decision time is similar to that in our baseline. In contrast, belief bias and variability loom larger in the "Short Histories" treatment, in line with suggestive evidence that these tend to be greater at histories leading to more moderate beliefs.

Garbled Information. Our last design variation (experiment B, "Garbled Info" treatment) considered the case in which subjects never perfectly learn θ. Our goal was to allow subjects to form non-degenerate beliefs about θ *even* following an observation of a truth ball. As before, when a draw s_t is labelled as noise ($n_t = 1$), it is an independently drawn uniform ϵ_t. Unlike our baseline, however, even when labelled as a truth ball ($n_t = 0$), s_t matches θ with 80% probability and is uniform noise with complementary probability.

The specific implementation of this design was as follows: At the start of each round, a *truth box* (instead of a truth ball) is chosen at random to be either "mostly yellow" or "mostly blue", each with equal probability. A "mostly yellow" box has 9 yellow balls and 1 blue ball, and vice versa for a "mostly blue" box. Subjects could observe draws from the truth box or from a *noise box* consisting of 5 yellow and 5 blue balls. For periods 1 and 2, a ball is drawn (with replacement) and shown to the subject; with probability 1/2, the ball is from the noise box, and with probability 1/2 the ball is from the truth box. In period 3, there is either a new draw or a 'fact-check' (a slight variation in terminology relative to 'validation' from the baseline design). In a fact-check, one of the prior draws is chosen uniformly at random, and the subject is told which box the ball is drawn from. In short, we simultaneously vary (i) the likelihood of new draws (from 3/2 to 7/3), (ii2) the probability of drawing a noise ball, and (iii) the fact that now observing a ball from the truth box does not fully reveal the urn composition.

Despite the changes to the design, our results stand. We again here find that subjects update less from retractions than from direct evidence, and behave as if it is more complex as per our

indicators: they take longer and exhibit greater belief bias and variability.[37]

7. Conclusions

This paper identifies and quantifies diminished updating from retractions and shows updating from retractions is revealed more complex. Our analyses distinguish diminished updating from retractions and other information processing patterns that may not have been previously recognized as relevant to retraction effectiveness. These findings provide insights into the design of interventions to address erroneous information. Specifically, we find that presenting direct evidence is more effective in correcting beliefs than retractions or corrections. Furthermore, corrections of erroneous evidence are more effective when they occur swiftly.

The minimality of our design facilitated a clear link between empirical results and their theoretical interpretation. But it certainly overlooks significant dimensions of real-world scenarios, where outcomes (e.g., citations) reflect factors other than probabilistic likelihood assessments, and domain-specific factors (e.g., memory frictions, motivated reasoning about health outcomes, etc.) may influence how individuals respond to retractions. However, the information structure in our study does approximate certain aspects of retractions in scientific articles, fact-checking, or other mechanisms of information correction. Furthermore, we interpret the consistency of our results across variations of our baseline design as evidence for the external validity of our mechanism. As such, our contribution is to propose that the additional layers of complexity in updating from retractions are generically an important factor in explaining the diminished updating from retractions. To the extent that other factors are significant, our work suggests their impacts should be separately identified from—and interacted with—the effects of retractions on information processing analyzed here.

Our results point to several interesting potential directions for future work. Two strike us as particularly natural.

First, studying what makes indirect information more complex. Our experiment was designed to highlight how errors in information processing contribute to retraction failures. The richness afforded to us by variation in the design spoke to our proposed mechanism without altering the fundamental nature of the task at hand. Our findings suggest scope to further elucidate patterns in cognitive noise in indirect information. In particular, our results point toward the need for theoretical models of costly information processing to distinguish direct from indirect

[37] Interestingly, they also update less from new draws—something in line with existing evidence that under-inference from evidence is higher the greater its likelihood (see Augenblick, Lazarus and Thaler, 2023).

information. Additional research is necessary to document how belief updating depends on the degree of contingent reasoning involved. This agenda is not only of theoretical interest but also practical importance, as it aims to clarify how to correct misinformation and improve information transmission.

Second, exploring the implications of these patterns on optimal information design policies. In many settings—e.g., interactions between politicians and the media, or firms and financial auditors—information receivers obtain results from strategic interplay between senders and third-party verification (e.g., Levkun, 2021). While our results suggest receivers may be susceptible to err following certain kinds of information, we do not speak to how endogenous changes in information may influence belief-updating patterns. Furthermore, a broader implication of our work is that the way in which information is generated can influence its perception beyond the objective informational content. While work in information design commonly reduces information to posterior beliefs, such reductions may omit important economic forces that seem worth exploring. For instance, knowing that retractions are not fully effective in correcting beliefs, to what extent could an information designer (e.g., a partisan media outlet or a political campaign strategist) exploit under-reaction to retractions? How would our findings shape their information policy, and how should a third party design a verification or fact-checking policy to counter it? If corrections but not validations are announced, will people correctly treat unretracted evidence as more reliable? When policies target evidence favoring a particular view, are the resulting corrections or fact-checks perceived as less informative? We believe answering these and related questions has substantial practical value.

8. References

Abdellaoui, M., P. Klibanoff, and L. Placido. 2015. "Experiments on Compound Risk in Relation to Simple Risk and to Ambiguity." *Management Science*, 61(6): 1306–1322.

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

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

Alaoui, L., K. Janezic, and A. Penta. 2020. "Reasoning about others' reasoning." *Journal of Economic Theory*, 189: 105091.

Ali, S. N., M. Mihm, L. Siga, and C. Tergiman. 2021. "Adverse and Advantageous Selection in the Laboratory." *American Economic Review*, 111(7): 2152–2178.

Alós-Ferrer, C., E. Fehr, and N. Netzer. 2021. "Time Will Tell: Recovering Preferences When Choices Are Noisy." *Journal of Political Economy*, 129(6): 1828–77.

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

Anderson, L. R., and C. A. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review*, 87(5): 847–62.

Angelucci, C., and A. Prat. 2020. "Measuring Voters' Knowledge of Political News." *Working Paper*.

Angrisani, M., A. Guarino, P. Jehiel, and T. Kitagawa. 2021. "Information Redundancy Neglect versus Overconfidence: A Social Learning Experiment." *American Economic Journal: Microeconomics*, 13(3): 163–97.

Augenblick, N., E. Lazarus, and M. Thaler. 2023. "Overinference from Weak Signals and Underinference from Strong Signals." *Working Paper*.

Azoulay, P., J. L. Furman, J. L. Krieger, and F. Murray. 2015. "Retractions." *Review of Economics and Statistics*, 97(5): 1118–1136.

Azrieli, Y., C. P. Chambers, and P. J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy*, 126(4): 1472–1503.

Ba, C., J. A. Bohren, and A. Imas. 2022. "Over and Underreaction to Information." *Working Paper*.

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(3): 435–447.

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.

Benjamin, D. 2019. "Errors in Probabilistic Reasoning and Judgment Biases." In *Handbook of Behavioral Economics.* , ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson. Elsevier Press.

Bhui, R., and S. J. Gershman. 2018. "Decision by Sampling Implements Efficient Coding of Psychoeconomic Functions." *Psychological Review*, 125(6): 985–1001.

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

Broockman, D., and J. Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science*, 352(6282): 220–224.

Buckley, P. B., and C. B. Gillman. 1974. "Comparison of Digits and Dot Patterns." *Journal of Experimental Psychology*, 103(6): 1131–1136.

Caplin, A., D. Csaba, J. Leahy, and O. Nov. 2020. "Rational Inattention, Competitive Supply, and Psychometrics." *Quarterly Journal of Economics*, 135(3): 1681–1724.

Chan, M.-p. S., C. R. Jones, K. H. Jamieson, and D. Albarracín. 2017. "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation." *Psychological Science*, 28(11): 1531–1546.

Charness, G., and C. Dave. 2017. "Confirmation bias with motivated beliefs." *Games and Eco-*

nomic Behavior, 104: 1–23.

Charness, G., and D. Levin. 2005. "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect." *American Economic Review*, 95(4): 1300–1309.

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(3): 1656–1691.

Coutts, A. 2019. "Good news and bad news are still news: experimental evidence on belief updating." *Experimental Economics*, 22: 369–395.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri. 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *Journal of Economic Literature*, 51(1): 5–62.

Cripps, M. 2021. "Divisible Updating." *Working Paper*.

Danz, D., L. Vesterlund, and A. J. Wilson. 2022. "Belief Elicitation and Behavioral Incentive Compatibility." *American Economic Review*, 112(9): 2851–83.

Dean, M., and P. Ortoleva. 2019. "The empirical relationship between nonstandard economic behaviors." *Proceedings of the National Academy of Sciences*, 116(33): 16262–16267.

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.

Ecker, U. K. H., Z. O'Reilly, J. S. Reid, and E. P. Chang. 2020. "The effectiveness of short-format refutational fact-checks." *British Journal of Psychology*, 111: 36–54.

Enke, B. 2020. "What You See is All There Is." *Quarterly Journal of Economics*, 135(3): 1363–1398.

Enke, B., and C. Shubatt. 2023. "Quantifying Lottery Choice Complexity." *Working Paper*.

Enke, B., and F. Zimmermann. 2019. "Correlation Neglect in Belief Formation." *Review of Economic Studies*, 86(1): 313–332.

Enke, B., and T. Graeber. 2021. "Cognitive uncertainty in intertemporal choice." *Working Paper*.

Enke, B., and T. Graeber. 2022. "Cognitive Uncertainty." *Quarterly Journal of Economics*, 138(4): 2021–2067.

Enke, B., T. Graeber, and R. Oprea. 2023*a*. "Complexity and Time." *Working Paper*.

Enke, B., T. Graeber, and R. Oprea. 2023*b*. "Confidence, Self-Selection, and Bias in the Aggregate." *American Economic Review*, 113(7): 1933–66.

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

Esponda, I., and E. Vespa. 2014. "Hypothetical Thinking and Information Extraction in the Laboratory." *American Economic Journal: Microeconomics*, 6(4): 180–202.

Esponda, I., and E. Vespa. 2021. "Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory." *Working Paper*.

Esponda, I., E. Vespa, and S. Yuksel. 2024. "Mental Models and Learning: The Case of Base-Rate Neglect." *American Economic Review*, 114(3): 752–82.

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

Eyster, E., M. Rabin, and D. Vayanos. 2019. "Financial Markets Where Traders Neglect the Informational Content of Prices." *Journal of Finance*, 74(1): 371–399.

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(42): 17028–17033.

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(11): 1215–1226.

Friedman, D. 1998. "Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly." *American Economic Review*, 88(4): 933–946.

Frydman, C., and L. J. Jin. 2022. "Efficient Coding and Risky Choice." *Quarterly Journal of Economics*, 137(1): 161–213.

Fudenberg, D., P. Strack, and T. Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review*, 108(12): 3651–84.

Gabis, L. V., O. L. Attia, M. Goldman, N. Barak, P. Tefera, S. Shefer, M. Shaham, and T. Lerman-Sagie. 2022. "The myth of vaccination and autism spectrum." *European Journal of Paediatric Neurology*, 36: 151–158.

Gonçalves, D. 2023. "Sequential Sampling Equilibrium." *Working Paper*.

Gonçalves, D. 2024. "Speed, Accuracy, and Complexity." *Working Paper*.

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*.

Greitemeyer, T. 2014. "Article retracted, but the message lives on." *Psychonomic Bulletin & Review*, 21: 557–561.

Grether, D. M. 1980. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *The Quarterly Journal of Economics*, 95(3): 537–557.

Guay, B., A. J. Berinsky, G. Pennycook, and D. G. Rand. 2023. "How to think about whether misinformation interventions work." *Nature Human Behavior*, 7: 1231–1233.

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

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

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya. 2023. "Curtailing False News, Amplifying Truth." *Working Paper*.

Halim, E., Y. E. Riyanto, and N. Roy. 2019. "Costly Information Acquisition, Social Networks, and Asset Prices: Experimental Evidence." *Journal of Finance*, 74(4): 1975–2010.

Harmon-Jones, E., and J. Mills. 2019. "An introduction to cognitive dissonance theory and an overview of current perspectives on the theory." In *Cognitive dissonance: Reexamining a pivotal theory in psychology.* , ed. Eddie Harmon-Jones, 3–24. American Psychological Association.

Hartzmark, S. M., S. D. Hirshman, and A. Imas. 2021. "Ownership, Learning, and Beliefs."

Quarterly Journal of Economics, 136(3): 1665–1717.

Healy, P. J., and J. Kagel. 2023. "Testing Elicitation Mechanisms via Team Chat." *Working Paper*.

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

Hussinger, K., and M. Pellens. 2019. "Guilt by association: How scientific misconduct harms prior collaborators." *Research Policy*, 48(2): 516–530.

Jacoby, L., C. Kelley, J. Brown, and J. Jasechko. 1989. "Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past." *Journal of Personality and Social Psychology*, 56(3): 326–338.

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(6): 1420–1436.

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(10): 1046–1054.

Khaw, M. W., Z. Li, and M. Woodford. 2021. "Cognitive Imprecision and Small-Stakes Risk Aversion." *Review of Economic Studies*, 88(4): 1979–2013.

Kneeland, T. 2015. "Identifying Higher-Order Rationality." *Econometrica*, 83(5): 2065–79.

Krajbich, I., C. Armel, and A. Rangel. 2010. "Visual fixations and the computation and comparison of value in simple choice." *Nature Neuroscience*, 13(10): 1292–1298.

Krajbich, I., D. Lu, C. Camerer, and A. Rangel. 2012. "The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions." *Frontiers in Psychology*, 3(193): 235–251.

Levkun, A. 2021. "Communication with Strategic Fact-checking." *Working Paper*.

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(3): 106–131.

Liang, Y. 2020. "Learning from unknown information sources." *Working Paper*.

Lu, S. F., G. Z. Jin, B. Uzzi, and B. Jones. 2013. "The Retraction Penalty: Evidence from the Web of Science." *Scientific Reports*, 3146(3): 1–5.

Martínez-Marquina, A., M. Niederle, and E. Vespa. 2019. "Failures in Contingent Reasoning: The Role of Uncertainty." *American Economic Review*, 109(10): 3437–3474.

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

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(3): 144–162.

Mobius, M. M., M. Niederle, P. Niehaus, and T. Rosenblat. 2022. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science*, 68(11): 7793–7817.

Motta, M., and D. Stecula. 2021. "Quantifying the effect of Wakefield et al. (1998) on skepticism about MMR vaccine safety in the U.S." *PLoS ONE*, 16(8): e0256395.

Nyhan, B. 2021. "Why the backfire effect does not explain the durability of political misperceptions." *Proceedings of the National Academy of Sciences*, 118(15).
Nyhan, B., and J. Reifler. 2010. "When corrections fail: The persistence of political misperceptions." *Political Behavior*, 32(2): 303–330.
Nyhan, B., E. Porter, J. Reifler, and T. Wood. 2019. "Taking Fact-checks Literally But Not Seriously? The Effects of Journalistic Fact-checking on Factual Beliefs and Candidate Favorability." *Political Behavior*, forthcoming.
Oprea, R. 2020. "What Makes a Rule Complex?" *American Economic Review*, 110(12): 3913–3951.
Oprea, R. 2022. "Simplicity Equivalents." *Working Paper*.
Oprea, R., and S. Yuksel. 2022. "Social Exchange of Motivated Beliefs." *Journal of the European Economic Association*, 20(2): 667–699.
Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press.
Pearl, J., and D. Mackenzie. 2018. *The Book of Why*. Basic Books.
Pennycook, G., and D. G. Rand. 2021. "The Psychology of Fake News." *Trends in Cognitive Sciences*, 1–29.
Pennycook, G., J. Binnendyk, C. Newton, and D. G. Rand. 2021. "A Practical Guide to Doing Behavioral Research on Fake News and Misinformation." *Collabra: Psychology*, 7.
Pluviano, S., C. Watt, and S. Della Sala. 2017. "Misinformation lingers in memory: Failure of three pro-vaccination strategies." *PLoS ONE*, 12(7): e0181640.
Pullan, S., and M. Dey. 2021. "Vaccine hesitancy and anti-vaccination in the time of COVID-19: A Google Trends analysis." *Vaccine*, 39(14): 1877–1881.
Rabin, M., and J. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 144(1): 37–82.
Ratcliff, R. 1978. "A Theory of Memory Retrieval." *Psychological Review*, 85(2): 59.
Ratcliff, R., P. L. Smith, S. D. Brown, and G. McKoon. 2016. "Diffusion Decision Model: Current Issues and History." *Trends in Cognitive Sciences*, 20(4): 260–281.
Retraction Watch. 2023. "Top 10 most highly cited retracted papers." *https://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/*, Accessed: Thursday 13th June, 2024.
Serra-Garcia, M., and U. Gneezy. 2021. "Nonreplicable publications are cited more than replicable ones." *Science Advances*, 7(21): 7.
Shishkin, D., and P. Ortoleva. 2021. "Ambiguous Information and Dilation: An Experiment." *Journal of Economic Theory*, Forthcoming.
Snowberg, E., and L. Yariv. 2021. "Testing the Waters: Behavior across Participant Pools." *American Economic Review*, 111(2): 687–719.
Susmann, M. W., and D. T. Wegener. 2022. "The role of discomfort in the continued influence effect of misinformation." *Memory and Cognition*, 50(2): 435–448.
Swire, B., A. J. Berinsky, S. Lewandowsky, and U. K. H. Ecker. 2017. "Processing political misinformation: comprehending the Trump phenomenon." *Royal Society of Open Science*, 4(3).

Taber, C. S., and M. Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science*, 50(3): 755–769.
Tan, H.-T., and S.-K. Tan. 2009. "Investors' reactions to management disclosure corrections: Does presentation format matter?" *Contemporary Accounting Research*, 26(2): 605–626.
Tan, S.-K., and L. Koonce. 2011. "Investors' reactions to retractions and corrections of management earnings forecasts." *Accounting, Organizations and Society*, 36(6): 382–397.
Thaler, M. 2020. "The "Fake News" Effect: Experimentally Identifying Motivated Reasoning Using Trust in News." *Working Paper*.
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.
Walter, N., and R. Tukachinsky. 2020. "A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?" *Communication Research*, 47(2): 155–177.
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.
Weizsäcker, G. 2010. "Do We Follow Others When We Should? A Simple Test of Rational Expectations." *American Economic Review*, 100(5): 2340–60.
Woodford, M. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics*, , (1): 579–601.
Wright, G., and P. Ayton. 1988. "Decision time, subjective probability, and task difficulty." *Memory & Cognition*, 16(2): 176–185.

Appendix.

Appendix A. Additional Discussion of the Related Literature

In this appendix, we discuss existing domain-specific evidence of retraction ineffectiveness.

Political Information. Perhaps the largest number of experiments in this literature have studied the correction of information in political settings. While interpreting magnitudes is sometimes difficult in these studies, most show retractions have diminished effectiveness in political contexts.[38] For instance, in the context of the 2016 US Presidential election (Swire et al., 2017; Nyhan et al., 2019) and the 2017 French Presidential election (Barrera et al., 2020), fact-checking did improve factual knowledge, but was less effective than the original corrected information. Guriev et al. (2023), however, document relatively small impacts of fact-checking on perceived veracity in the context of the 2022 US Midterm elections. Many studies suggest motivated reasoning as the main explanation for the ineffectiveness of retractions in political contexts.[39] Although it may indeed play a significant role, our results indicate that retractions fail even in the absence of motivated reasoning.

Fake News. Prior literature on fake news across psychology, political science and economics has studied the effectiveness of fact checking in combating misinformation; Pennycook and Rand (2021) discuss several reasons for this apparent diminished effectiveness. It is worth emphasizing that many papers in this literature vary the *nature of the fact-check itself*, with the pattern of interest being whether some presentations of fact-checks are viewed as subjectively more informative; see (Ecker et al., 2020) for both an insightful discussion and an example.

Financial Information. Other work has focused on the effectiveness of retractions in financial settings, where designs tend to involve presentations of earnings reports or related financial statements and then instructions to disregard. The focus is typically less on beliefs themselves, but rather how the information is *used* in assessments or investments. Grant, Hodge and Seto (2021), Tan and Tan (2009), and Tan and Koonce (2011) run experiments using such designs, finding that retractions have diminished effectiveness in these domains, and discuss ways this can be combated.

[38] In the context of highly politically charged topics, retractions may in rare cases *backfire*, leading subjects to believe more strongly in the retracted information. Nyhan and Reifler (2010) noted the occurrence of backfiring in an experiment where they provided subjects with information about the presence of weapons of mass destruction in Iraq during the early 2000s, and subsequently provided them with corrections. This extreme form of retraction failure, for the most part, has not been replicated. See Nyhan (2021) for an authoritative discussion.

[39] Various studies have articulated how motivated reasoning influences belief processing in political domains; for instance, see Angelucci and Prat (2020), Thaler (2020), and Taber and Lodge (2006).

Jury Trials. Jury trials often feature information which jurors are instructed to disregard. Experiments on this question tend to focus on whether the reason evidence should be disregarded matters. Kassin and Sommers (1997), Thompson, Fong and Rosenhan (1981) and Fein, McCloskey and Tomlinson (1997) conduct experiments documenting that juries do not always simply disregard information if instructed to do so. While these studies do show retracted information is not so easily disregarded, it is less clear that this reflects a departure from Bayesian rationality, since the retracted information is often meaningful.

Academic Papers. In addition to work studying society's beliefs in the association between vaccines and autism discussed in the introduction, other existing literature on retractions of scientific articles typically focuses on documenting the reasons why papers are retracted, as well as assessing the consequences for researchers. While fraud and academic misconduct are the main reasons behind retractions, error and failure to replicate constitute a significant fraction of the retraction notices (Brainard and You, 2018; Fang, Steen and Casadevall, 2012)—and it is important to note that many papers that do not replicate are not retracted (Serra-Garcia and Gneezy, 2021). Among the academic community, there seems to be a significant penalty for researchers associated to retractions: a decrease in citations not only of the authors' prior work, but also of their collaborators', and, more generally, of work in related topics (Lu et al., 2013; Azoulay et al., 2015; Hussinger and Pellens, 2019). Of course, there are many reasons citations may be an imperfect proxy for retraction effectiveness (strategic citation motives, information about retractions not reaching the target audience, among others). Existing experimental evidence focusing on this setting suggests that retractions induce insufficient belief updating, even when the cited reason is fabricated data, and points to availability of a causal narrative as a possible reason (see, e.g., Greitemeyer, 2014). While these studies do show that retracted information is not so easily disregarded, relying on observational data is challenging. Indeed at least in some cases, the diminished updating from retractions may not reflect a misperception of its informational content and instead be consistent with Bayesian inference; for instance, a scientific article's retraction may involve a dispute with unclear implications, or follow-on work may find the retracted article made certain contributions which were accepted as valid (see Fang, Steen and Casadevall (2012) for examples).

Appendix B. Sample Characteristics

	MTurk	Elicit at End	Garbled Info	Prolific	No History	Retraction Info	Short Histories
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total Subjects	211	204	164	155	164	164	150
Age	37.7	39.5	38.9	38.5	37.8	37.6	35.5
Female	0.398	0.402	0.433	0.477	0.439	0.500	0.473
High School	0.900	0.887	0.927	0.865	0.860	0.884	0.880
College Degree	0.673	0.613	0.683	0.587	0.506	0.561	0.560
Postgraduate	0.180	0.201	0.189	0.148	0.177	0.177	0.153
High Comprehension	0.602	0.495	0.561	0.574	0.634	0.555	0.660
High Quant	0.275	0.294	0.171	0.290	0.317	0.317	0.307
Experiment	A	A	B	C	C	C	D
Date	2020-06	2020-06	2021-05	2024-01	2024-01	2024-01	2024-02
Platform	MTurk	MTurk	MTurk	Prolific	Prolific	Prolific	Prolific

Table 5: Sample Characteristics

Notes: The table shows sample characteristics for each of our treatments in each of our experiments. "Age" is measured in years; "Female" denotes the fraction of the sample that identifies as a woman; "High School", "College Degree", and "Postgraduate Studies" denote the fraction of the sample that has completed the respective level of education. "Comprehension Correct" shows the fraction of the sample that answered all comprehension questions correctly at first try; "High Quant" shows to the fraction of subjects who answer all the quantitative questions correctly at on their first try. Finally, "Date" denotes when the data was collected, and "Platform" the venue used to recruit subjects.

Appendix C. Comparison of Recruitment Platforms

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(\mathrm{T}_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(\mathrm{T}_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.009	2.454***	0.061***	1.146***	-0.031	0.797**	0.080***	0.486*
	(0.027)	(0.306)	(0.016)	(0.232)	(0.030)	(0.331)	(0.016)	(0.251)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.600***	–	–	–	0.608***	–	–	–
	(0.090)				(0.104)			
Retraction (r_t) x MTurk	0.004	0.544	0.005	0.163	0.021	0.544	0.007	0.164
	(0.035)	(0.464)	(0.021)	(0.275)	(0.034)	(0.464)	(0.020)	(0.276)
Retracted Draw x MTurk	-0.024	–	–	–	-0.007	–	–	–
	(0.131)				(0.131)			
Mean Decision Time			8.830				8.830	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.27	0.08	0.18	0.03	0.27	0.08	0.18	0.03
N	39162	39162	39162	5236	39162	39162	39162	5236

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

Table 6: Treatment Effects across Recruitment Platform

Notes: This table compares average treatment effects in our baseline treatment across experiments A (MTurk) and C (Prolific). MTurk corresponds to an indicator variable that equals 1 when the observation is from our baseline treatment in experiment A. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline treatments, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

Appendix D. Comprehension Questionnaire

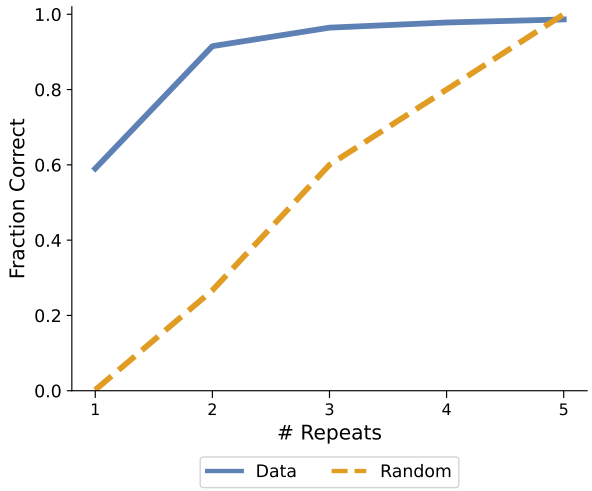
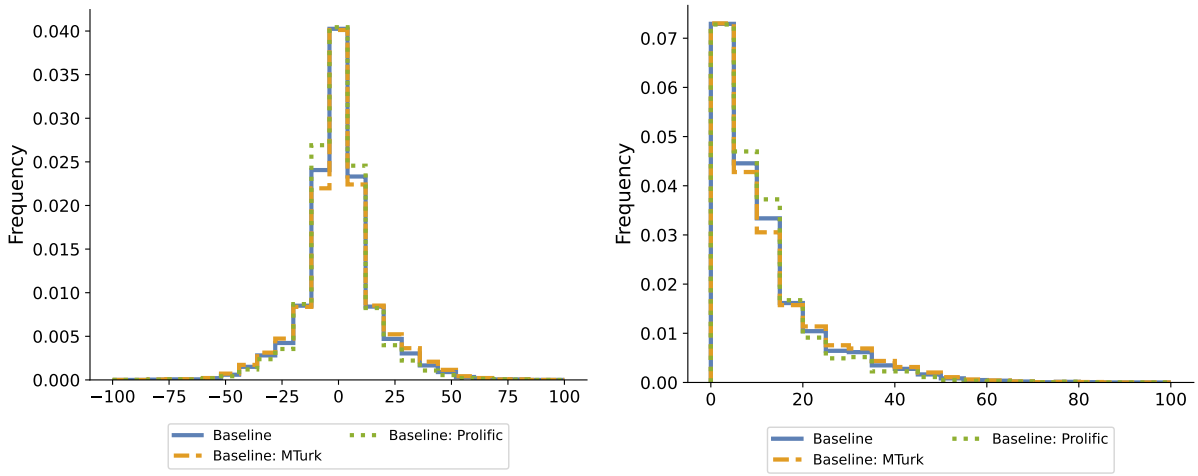


Figure 10: Comprehension Questions

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.

Online Appendix.

Online Appendix A. Comparing Beliefs to the Bayesian Posterior



(a) Difference to Bayesian Posteriors

(b) Distance from Bayesian Posteriors

Figure 11: Comparing Beliefs to Bayesian Posteriors

Notes: The figure shows (a) the average difference between beliefs and the Bayesian posterior and (b) the average distance between the beliefs and the Bayesian posterior. The sample comprises the baseline treatments in both experiments A (MTurk; dashed orange line) and C (Prolific; dotted green line), as well as the pooled sample (solid blue line), and it includes all periods in which the history only include new draws.

Online Appendix B. Retraction Ineffectiveness

In this section, we present results from Section 3 fully disaggregated by history. In Section B.1 we report beliefs (in levels) by history, first at histories that only include new draws, and then at histories in which the last observation is a retraction. In Section B.2, we report treatment effects (in log-odds) of retraction ineffectiveness, corresponding to Table 2, disaggregated by sign history for comparisons to no retracted draw (test (a)), and by compressed history for comparisons to equivalent new draws (test (b)).

B.1. Beliefs Disaggregated by History

B.1.1. Beliefs Following New Draws

Sign History (1)	Bayesian Posterior (2)	Mean Reported Belief (3)	Obs (4)
$BBBB$	16.495	20.750 (1.773)	208
BBB	22.857	26.279 (0.865)	785
BB	30.769	35.435 (0.301)	3027
$BBBY$	30.769	28.268 (1.634)	194
$BBYB$	30.769	28.264 (1.708)	170
$YBBB$	30.769	31.357 (1.939)	173
$BYBB$	30.769	23.832 (1.498)	165
B	40.000	42.904 (0.188)	5771
BBY	40.000	40.982 (0.629)	691
BYB	40.000	38.142 (0.530)	662
YBB	40.000	43.105 (0.655)	685
$BBYY$	50.000	47.997 (1.117)	171
BY	50.000	50.127 (0.235)	2744
$BYBY$	50.000	48.888 (0.979)	143
$BYYB$	50.000	50.496 (0.926)	154
YB	50.000	51.800 (0.226)	2765
$YBBY$	50.000	50.232 (0.914)	176
$YBYB$	50.000	54.502 (1.146)	163
$YYBB$	50.000	53.827 (0.976)	196
BYY	60.000	61.127 (0.564)	677
Y	60.000	57.627 (0.173)	5941
YBY	60.000	61.725 (0.514)	677
YYB	60.000	58.987 (0.530)	760
$BYYY$	69.231	75.486 (1.598)	183
$YBYY$	69.231	74.454 (1.554)	164
YY	69.231	63.855 (0.284)	3176
$YYBY$	69.231	70.677 (1.714)	172
$YYYB$	69.231	72.480 (1.618)	212
YYY	77.143	75.432 (0.757)	829
$YYYY$	83.505	80.039 (1.545)	230

Standard errors in parentheses.

Table 7: Beliefs Disaggregated by History: New Draws

Notes: The table shows, (1) for each history of draws, (2) the associated Bayesian Posterior, (3) the average beliefs for our baseline treatment (pooling experiments A and C), and (4) the number of observations. The sample only includes periods t in which the history $\mathscr{H}_t$ only includes new draws.

B.1.2. Beliefs Following Retractions

Sign History (1)	Bayesian Posterior (2)	Mean Reported Belief (3)	Obs (4)
$BBBY$	30.769	27.219 (1.401)	297
$BBYB$	30.769	31.651 (1.770)	92
$YBBB$	30.769	37.160 (2.138)	95
$BYBB$	30.769	37.161 (2.049)	89
BBY	40.000	38.829 (0.490)	1176
BYB	40.000	42.965 (0.613)	566
YBB	40.000	47.019 (0.594)	607
$BBYY$	50.000	42.871 (1.124)	199
$BYBY$	50.000	44.870 (1.141)	198
$BYYB$	50.000	57.404 (1.261)	178
$YBBY$	50.000	48.191 (1.478)	184
$YBYB$	50.000	57.360 (0.923)	193
$YYBB$	50.000	57.132 (0.921)	207
BYY	60.000	55.819 (0.629)	599
YBY	60.000	58.012 (0.574)	569
YYB	60.000	60.423 (0.464)	1242
$BYYY$	69.231	66.010 (1.806)	98
$YBYY$	69.231	65.646 (1.485)	102
$YYBY$	69.231	67.820 (1.701)	112
$YYYB$	69.231	73.608 (1.334)	295

Standard errors in parentheses.

Table 8: Beliefs Disaggregated by History: Retractions

Notes: The table shows, (1) for each history of draws, (2) the associated Bayesian Posterior, (3) the average beliefs for our baseline treatment (pooling experiments A and C), and (4) the number of observations. The sample only includes periods t in which the lagged history $\mathscr{H}_{t-1}$ only includes new draws and a retraction is observed in period t.

B.2. Retraction Ineffectiveness: Treatment Effects by History

Comparison to Retractions	Sign/Compressed History	Signal	Obs	ATE
(1)	(2)	(3)	(4)	(5)
Equivalent New Draw	$BBBY$	+1/Y	491	0.462 (0.627)
Equivalent New Draw	BBY	+1/Y	1867	0.326 (0.161)
Equivalent New Draw	$BBYB$	-1/B	262	1.524 (0.668)
Equivalent New Draw	$BBYY$	+1/Y	370	0.515 (0.261)
Equivalent New Draw	BYB	-1/B	1228	0.526 (0.210)
Equivalent New Draw	$BYBB$	-1/B	254	3.344 (0.661)
Equivalent New Draw	$BYBY$	+1/Y	341	0.666 (0.298)
Equivalent New Draw	BYY	+1/Y	1276	0.675 (0.168)
Equivalent New Draw	$BYYB$	-1/B	332	1.085 (0.244)
Equivalent New Draw	$BYYY$	+1/Y	281	2.921 (0.707)
Equivalent New Draw	YBB	-1/B	1292	0.401 (0.158)
Equivalent New Draw	$YBBB$	-1/B	268	1.947 (0.678)
Equivalent New Draw	$YBBY$	+1/Y	360	0.331 (0.344)
Equivalent New Draw	YBY	+1/Y	1246	0.384 (0.129)
Equivalent New Draw	$YBYB$	-1/B	356	0.155 (0.247)
Equivalent New Draw	$YBYY$	+1/Y	266	2.314 (0.689)
Equivalent New Draw	YYB	-1/B	2002	0.178 (0.161)
Equivalent New Draw	$YYBB$	-1/B	403	0.363 (0.238)
Equivalent New Draw	$YYBY$	+1/Y	284	1.034 (0.606)
Equivalent New Draw	$YYYB$	-1/B	507	0.121 (0.604)
No Retracted Draw	B	-1/B	6944	0.311 (0.104)
No Retracted Draw	B	+1/Y	6947	0.585 (0.101)
No Retracted Draw	BB	-1/B	3303	-0.212 (0.237)
No Retracted Draw	BB	+1/Y	3324	3.033 (0.421)
No Retracted Draw	BY	-1/B	3019	0.938 (0.186)
No Retracted Draw	BY	+1/Y	3043	1.010 (0.191)
No Retracted Draw	Y	-1/B	7183	0.373 (0.100)
No Retracted Draw	Y	+1/Y	7109	0.112 (0.116)
No Retracted Draw	YB	-1/B	3068	0.618 (0.130)
No Retracted Draw	YB	+1/Y	3047	0.684 (0.196)
No Retracted Draw	YY	-1/B	3471	3.247 (0.495)
No Retracted Draw	YY	+1/Y	3488	-0.717 (0.242)

Clustered standard errors at the subject level in parentheses.

Table 9: Updating from Retractions: Disaggregated by History

Notes: The table shows, the average treatment effect of a retraction on beliefs in log-odds ($\hat{\ell}_t$) for our baseline treatment (pooling experiments A and C) as estimated in Table 2, but disaggregated by compressed history ($C(\mathcal{H}_t)$) when comparing beliefs with a retraction and without the retracted observation (No Retracted Draw), or by sign history ($S(\mathcal{H}_t)$) when comparing beliefs with a retraction and with an equivalent new observation (Equivalent New Draw). Column (1) determines the comparison ('No Retracted Draw' or 'Equivalent New Draw'), Column (2) the sign or compressed history, Column (3) the signal implied by the draw or the retraction, Column (4) the number of observations, and Column (5) the average treatment effect (ATE) with clustered standard errors in parentheses. As in Table 2, the sample includes beliefs in log-odds ($\hat{\ell}_t$) of subjects in the baseline treatments, excluding cases in which the truth ball is disclosed and in which there was a retraction in the past.

Online Appendix C. Informational Complexity and Retraction Ineffectiveness

This section supports Section 4 of the paper. In Section C.1, we provide a short proof validating our measure of complexity based on the variability of belief reports, as referenced in Section 4.2. In Section C.2, we present tables corresponding to Figure 5 of Section 4.4.

C.1. Conditions for Higher Variability with Higher Complexity

We recall that, in our model in Section 4.2, posterior log-odds updates are given by

$$\hat{\ell}_t = \hat{\ell}_{t-1} + \beta K(E) + \beta \frac{\sigma_\zeta}{\sqrt{T}} \zeta,$$

with $\beta = (1+\sigma_\zeta^2/(\sigma^2 T))^{-1}$ and $\zeta \sim \mathcal{N}(0,1)$. Further note that $\frac{\partial}{\partial \sigma_\zeta^2}\beta < 0$ implies that $T'(\sigma_\zeta^2)\frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)} < 1$, i.e., that the elasticity of T with respect to σ_ζ^2 is lower than 1. From the prediction that decision times increase with complexity, $T'(\sigma_\zeta^2) > 0$, we obtain that T is inelastic with respect to σ_ζ^2. Then, given $\hat{\ell}_{t-1}$ and $K(E)$, the variance of the posterior log-odds beliefs is

$$\mathrm{Var}(\hat{\ell}_t) = \beta^2 \frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)} = \sigma^2 \frac{\frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)}}{\left(\frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)} + \sigma^2\right)^2}.$$

Hence,

$$\frac{\partial}{\partial \sigma_\zeta^2}\mathrm{Var}(\hat{\ell}_t) = \frac{\sigma^2}{T(\sigma_\zeta^2)\frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)} + \sigma^2}^3 \left(\sigma^2 - \frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)}\right)\left(1 - T'(\sigma_\zeta^2)\frac{\sigma_\zeta^2}{T(\sigma_\zeta^2)}\right) > 0 \Longrightarrow \sigma^2/2 > (T(\sigma_\zeta^2)\sigma_\zeta^{-2} + \sigma^{-2})^{-1}.$$

C.2. Validating and Varying Complexity

We present tables for Figure 5.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.006	2.244***	0.038**	0.676***	0.014	1.286***	0.065***	0.288
	(0.027)	(0.284)	(0.015)	(0.239)	(0.038)	(0.336)	(0.018)	(0.223)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.392***	–	–	–	0.432***	–	–	–
	(0.083)				(0.096)			
Mean Decision Time			8.774				8.774	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.27	0.08	0.01	0.03	0.28	0.08	0.01	0.03
N	35211	35211	35211	4643	35211	35211	35211	4643

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

Table 10: Treatment Effects: Retract Last Draw

Notes: This table reports the effect of retractions on updating and empirical complexity measures when the last ball is retracted, corresponding to "Retract Last Draw" of Figure 5. Columns (1) and (5) are the regressions from Table 2, but restricted to retractions of the last draw. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	Equivalent New Draw			
	(1)	(2)	(3)	(4)
	$\hat{\ell}_t$	$\lvert\hat{p}_t - p_t\rvert$	$\ln(\mathrm{T}_t)$	$\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_{t-1})	0.026	2.270***	0.082***	0.803**
	(0.057)	(0.455)	(0.021)	(0.390)
Retraction × Signal ($r_{t-1} \cdot K(s_t)$)	0.751***	–	–	–
	(0.162)			
Mean Decision Time			8.794	
Sign History FEs	Yes	Yes	Yes	Yes
R^2	0.26	0.08	0.01	0.03
N	39168	39168	39168	4494

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

Table 11: Treatment Effects: After Retractions

Notes: This table reports the effect of retractions on updating from subsequent signals, corresponding to "After Retractions" of Figure 5.

We also exhibit the heterogeneous treatment effects for the retraction of the last draw, compared to retracting an earlier one.

Retraction vs.	No Retracted Draw			
	(1)	(2)	(3)	(4)
	$\hat{\ell}_t$	$\lvert\hat{p}_t - p_t\rvert$	$\ln(T_t)$	$\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.014	3.098***	0.080***	1.030***
	(0.024)	(0.313)	(0.014)	(0.168)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.740***	–	–	–
	(0.079)			
Retraction of Last Draw ($1\{\rho_t = t-1\}$)	-0.008	-0.760***	-0.037**	-0.231
	(0.037)	(0.284)	(0.017)	(0.219)
Retraction of Last Draw x Retracted Draw ($1\{\rho_t = t-1\}K(s_\rho)$)	-0.348***	–	–	–
	(0.091)			
Mean Decision Time			8.830	
Compressed History FEs	Yes	Yes	Yes	Yes
R^2	0.27	0.08	0.01	0.02
N	39162	39162	39162	5709

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

Table 12: Heterogeneous Treatment Effects: Retract Last Draw

Notes: This table reports heterogeneous treatment effects of retractions on updating and empirical complexity measures, based on whether the last ball is retracted. While Table 10 considers the treatment effect restricted to when the last draw is retracted, here we use the full baseline sample and report treatment effect heterogeneity by it.

Online Appendix D. Belief Updating Patterns under Retractions

This section supports Section 5 of the paper. In Section D.1, we present the results from Table 4, broken down by platform (MTurk vs. Prolific). In Section D.2, we examine heterogeneous treatment effects on our measures of complexity across histories, following the same disaggregation as in Figure 6.

D.1. Regression Tables

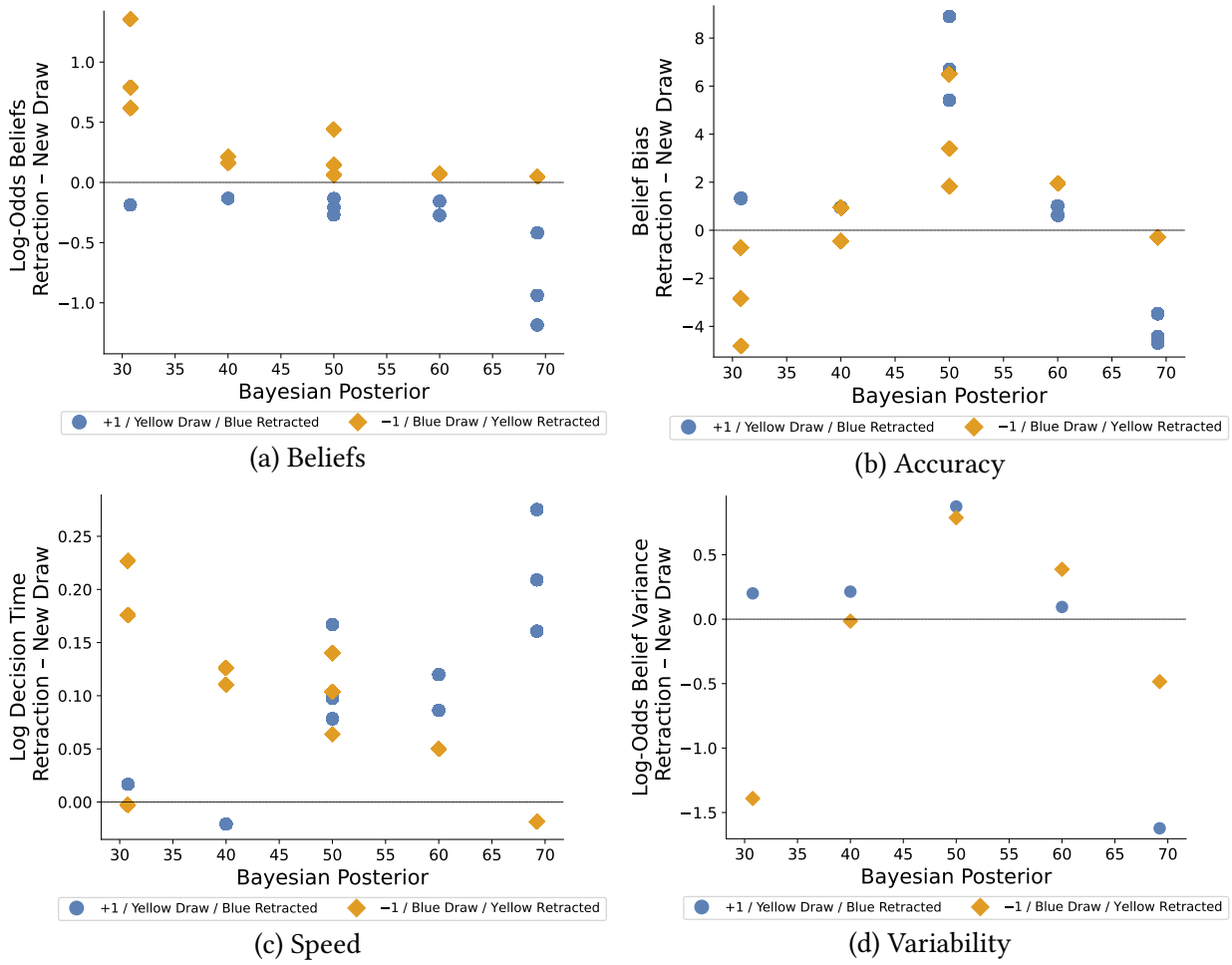
	Baseline		MTurk		Prolific	
	(1) $\hat{\ell}_t$	(2) $\hat{\ell}_t$	(3) $\hat{\ell}_t$	(4) $\hat{\ell}_t$	(5) $\hat{\ell}_t$	(6) $\hat{\ell}_t$
Signal (K_t)	1.102^{***}	0.907^{***}	1.126^{***}	0.998^{***}	1.066^{***}	0.788^{***}
	(0.060)	(0.060)	(0.071)	(0.072)	(0.102)	(0.102)
Prior (l_{t-1})	0.801^{***}	0.747^{***}	0.834^{***}	0.800^{***}	0.742^{***}	0.664^{***}
	(0.032)	(0.032)	(0.037)	(0.037)	(0.052)	(0.044)
Confirmatory Signal ($K_t \cdot c_t$)	–	0.650^{***}	–	0.417^{***}	–	0.958^{***}
		(0.097)		(0.135)		(0.128)
R^2	0.38	0.38	0.41	0.41	0.33	0.34
N	32064	32064	18491	18491	13573	13573

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

Table 13: Robustness of Grether Regressions by Experimental Platform

Notes: This table disaggregates the results from Table 4 by experimental platform.

D.2. Current Version/Figures



(a) Beliefs

(b) Accuracy

(c) Speed

(d) Variability

Figure 12: Updating from Retractions: Heterogeneity by History

Notes: This figure compares updating from retractions versus equivalent new draws, our test (b), disaggregated by sign history (i.e. corresponding to the disaggregation in Figure 6). Blue circles represent sign histories in which the last signal was a retraction of a blue draw, or a new yellow draw. Orange diamonds represent those in which the last signal was a retraction of a yellow draw, or a new blue draw. In both cases, the x-axis is the Bayesian posterior of the sign history. Panel (a) displays the effect of retractions on belief updating, $\hat{\ell}_t$. Panels (b)-(d) display effects on our three complexity indicators—accuracy ($|\hat{p}_t - p_t|$), speed ($\ln(T_t)$), and variability ($\text{Var}(\hat{\ell}_t \mid h_t)$). The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

Online Appendix E. Robustness 1: Subject Screening and Understanding

This section supports Section 6.1 of the paper. In Section E.1, we report regression tables corresponding to Figure 7. In Section E.2 we provide how confidence relates to the strength of updating to the kind of information subjects are presented with. In Section E.3, we disclose the estimates for heterogeneous treatment effects regarding the provision of additional information about retractions, comparing our baseline treatment and the "Retraction Information" treatment from experiment C.

E.1. Regression Tables Corresponding to Figure 7

We include a table for each of the robustness checks presented in the figure, each of which presents our main results on updating and empirical complexity measures but restricting the sample considering various dimensions of demonstrated understanding.

E.1.1. Comprehension Questionnaire Correct at 1st Try

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.004	2.635***	0.068***	1.014***	-0.036*	1.643***	0.090***	0.638***
	(0.016)	(0.320)	(0.015)	(0.175)	(0.022)	(0.340)	(0.017)	(0.190)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.596***	–	–	–	0.642***	–	–	–
	(0.076)				(0.086)			
Mean Decision Time			8.688				8.688	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.33	0.09	0.01	0.05	0.34	0.09	0.01	0.06
N	23110	23110	23110	3079	23110	23110	23110	3079

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

Table 14: Treatment Effects: Comprehension Correct

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects who answered all experimental comprehension questions, corresponding to "Comprehension Correct" of Figure 7. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

E.1.2. Understands Disclosure of Truth Ball

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.007	2.450***	0.078***	0.856***	-0.034	1.798***	0.098***	0.458**
	(0.017)	(0.355)	(0.017)	(0.191)	(0.024)	(0.379)	(0.020)	(0.214)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.541***	–	–	–	0.552***	–	–	–
	(0.079)				(0.088)			
Mean Decision Time			8.413				8.413	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.39	0.11	0.01	0.04	0.40	0.11	0.02	0.05
N	17958	17958	17958	2402	17958	17958	17958	2402

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

Table 15: Treatment Effects: Understand Disclosure

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects who, when the state is revealed, correctly report that they know the state, corresponding to "Understand Disclosure" of Figure 7. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

E.1.3. Few Mistakes

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.044**	2.981***	0.072***	0.944***	-0.047*	1.650***	0.100***	0.708***
	(0.018)	(0.392)	(0.018)	(0.190)	(0.027)	(0.394)	(0.021)	(0.191)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.674***	–	–	–	0.763***	–	–	–
	(0.090)				(0.108)			
Mean Decision Time			8.449				8.449	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.44	0.13	0.01	0.04	0.44	0.14	0.02	0.04
N	17357	17357	17357	2332	17357	17357	17357	2332

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

Table 16: Treatment Effects: Few Mistakes

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects who update in the opposite direction to the signal more than 10% of the time, corresponding to "Few Mistakes" of Figure 7. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

E.1.4. Understands Replacement

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.000	2.551***	0.057***	0.964***	-0.027	1.311***	0.079***	0.405***
	(0.017)	(0.268)	(0.012)	(0.164)	(0.025)	(0.299)	(0.014)	(0.155)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.482***	–	–	–	0.507***	–	–	–
	(0.063)				(0.085)			
Mean Decision Time			8.925				8.925	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.23	0.06	0.01	0.03	0.24	0.07	0.01	0.04
N	36059	36059	36059	4820	36059	36059	36059	4820

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

Table 17: Treatment Effects: Understands Replacement

Notes: This table reports the effect of retractions on updating and empirical complexity measures when excluding subjects from the baseline sample excludes subjects who could be mistaking sampling with and without replacement, corresponding to "Understands Replacement" of Figure 7. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3

E.1.5. Retraction Information

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.002	1.731***	0.098***	0.752***	0.015	0.625	0.087***	0.157
	(0.022)	(0.344)	(0.017)	(0.199)	(0.030)	(0.386)	(0.018)	(0.234)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.580***	–	–	–	0.856***	–	–	–
	(0.091)				(0.121)			
Mean Decision Time			12.256				12.256	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.30	0.09	0.01	0.04	0.29	0.09	0.01	0.04
N	17553	17553	17553	2338	17553	17553	17553	2338

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

Table 18: Treatment Effects: Understands Replacement

Notes: This table reports the effect of retractions on updating and empirical complexity measures in the treatment conveying additional retraction information in experiment C, corresponding to "Retraction Info" of Figure 7. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3

E.2. Subject Confidence

	(1) $\hat{\ell}_t$	(2) $\hat{\ell}_t$
Signal ($K(s_t)$)	1.008***	0.801***
	(0.172)	(0.186)
Prior (l_{t-1})	0.672***	0.620***
	(0.060)	(0.047)
Confirmatory Signal ($K(s_t) \cdot c_t$)	–	0.700***
		(0.171)
High Confidence x Signal ($K(s_t)$)	0.114	-0.023
	(0.200)	(0.213)
High Confidence x Prior (l_{t-1})	0.153*	0.097
	(0.082)	(0.075)
High Confidence x Confirmatory Signal ($K(s_t) \cdot c_t$)	–	0.513*
		(0.266)
R^2	0.33	0.34
N	13573	13573

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

Table 19: Belief Updating Patterns and Between-Subject Confidence: Grether Regressions

Notes: This table examines how between-subject heterogeneity relates to patterns in belief updating from new draws. It reports estimates of Equation 5 interacting the independent variables with whether or not the subject reports, on average, a higher confidence level than that of the median subject. The sample includes all observations of subjects in the baseline treatment in experiment C, in which we collect such confidence level measure, excluding periods in which the truth ball is disclosed, a retraction occurs, or in which there was a retraction in an earlier period.

In our experiment C, in each period we elicit how confident subjects are about their answer, as described in Section 6.1.

In Table 19, we estimate the standard Grether (1980) log-odds regression as per Equation 4, interacted with an indicator variable 'High Confidence' that equals 1 for subjects whose average reported confidence level is higher than that of the median subject. This corresponds to a between-subject analysis of how confidence relates to the belief updating patterns. We find that subjects who are more confident tend to update more, but especially so from confirmatory signals. They also tend to exhibit lower base-rate neglect. However, these patterns are not clear—they are, at best, significantly at a 10% significance level.

Table 20 exhibits the same regression but now interacted with the measure of confidence, 'Confidence', standardized within-subject. That is, for each subject, we subtract to the measure the mean reported confidence for that subject, and divide by the within-subject standard deviation of their reported level of confidence. We find beliefs react more strongly to information when subjects are more confident, but this is especially true for confirmatory information. Furthermore,

	(1) $\hat{\ell}_t$	(2) $\hat{\ell}_t$
Signal ($K(s_t)$)	1.097^{***}	0.771^{***}
	(0.101)	(0.110)
Prior (l_{t-1})	0.656^{***}	0.582^{***}
	(0.033)	(0.032)
Confirmatory Signal ($K(s_t) \cdot c_t$)	–	0.999^{***}
		(0.141)
Confidence x Signal ($K(s_t)$)	0.506^{***}	0.106^{*}
	(0.088)	(0.060)
Confidence x Prior (l_{t-1})	0.176^{***}	0.102^{***}
	(0.027)	(0.028)
Confidence x Confirmatory Signal ($K(s_t) \cdot c_t$)	–	1.186^{***}
		(0.184)
R^2	0.36	0.39
N	13316	13316

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

Table 20: Belief Updating Patterns and Within-Subject Confidence: Grether Regressions

Notes: This table examines how within-subject heterogeneity relates to patterns in belief updating from new draws. It reports estimates of Equation 5 interacting the independent variables with reported confidence, standardized within-subject. The sample includes all observations of subjects in the baseline treatment in experiment C, in which we collect such confidence level measure, excluding periods in which the truth ball is disclosed, a retraction occurs, or in which there was a retraction in an earlier period.

base-rate neglect is attenuated in instances in which subjects are more confident.

Table 21 examines if retractions affect confidence levels. It shows that indeed confidence is lower when subjects update from a retraction, but this effect is small and not robustly significant.

Retraction vs.	No Retracted Draw		Equivalent New Draw	
	(1)	(2)	(3)	(4)
	Confidence (Levels)	Confidence (Standardized)	Confidence (Levels)	Confidence (Standardized)
Retraction (r_t)	-0.878	-0.045	-1.928***	-0.091***
	(0.988)	(0.047)	(0.675)	(0.031)
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R^2	0.02	0.04	0.02	0.04
N	16267	16267	16267	16267

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

Table 21: Effect of Retractions on Confidence

Notes: This table provides estimates of the effect of retractions on confidence, following Equation 3. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(2)) and (b) updating from a retraction vs. an equivalent new draw (Columns (3)-(4)). Columns (1) and (3) refer to reported confidence levels, while Columns (2) and (4) use confidence levels standardized for each subject. The sample includes all observations of subjects in the baseline treatment in experiment C, in which we collect such confidence level measure, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

E.3. Additional Retraction Information

Retraction vs.	No Retracted Draw		Equivalent New Draw	
	(1)	(2)	(3)	(4)
	Confidence (Levels)	Confidence (Standardized)	Confidence (Levels)	Confidence (Standardized)
Retraction (r_t)	-0.840	-0.044	-2.061***	-0.102***
	(0.875)	(0.041)	(0.713)	(0.032)
Retraction Info	1.590	-0.005	1.602	-0.004
	(2.246)	(0.016)	(2.247)	(0.016)
Retraction Info x Retraction (r_t)	0.797	0.021	0.760	0.019
	(0.932)	(0.045)	(0.927)	(0.045)
Compressed History FEs	Yes	Yes	No	No
Sign History FEs	No	No	Yes	Yes
R^2	0.02	0.04	0.02	0.04
N	33820	33820	33820	33820

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

Table 22: Effect of Retractions on Confidence: Additional Retraction Information

Notes: This table provides estimates on heterogeneity regarding the effect of retractions on confidence, following Equation 3, interacting the main explanatory variable with an indicator, "Retraction Info", which equals 1 for subjects assigned to this treatment and is 0 for subjects assigned to our baseline treatment. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(2)) and (b) updating from a retraction vs. an equivalent new draw (Columns (3)-(4)). Columns (1) and (3) refer to reported confidence levels, while Columns (2) and (4) use confidence levels standardized for each subject. The sample includes all observations of subjects in the baseline and "Retraction Info" treatments in experiment C, in which we collect such confidence level measure, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

Table 22 shows that providing additional information about retractions leads to subjects reporting higher confidence levels when updating from retractions, even if only marginally so and not in a statistically significant manner.

Table 23 reports on heterogeneous treatment effects regarding the provision of additional information about retractions. No significant differences are detected.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.007	2.271***	0.075***	1.173***	-0.018	0.934***	0.063***	0.542**
	(0.027)	(0.304)	(0.016)	(0.232)	(0.030)	(0.321)	(0.017)	(0.241)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.602***	–	–	–	0.751***	–	–	–
	(0.090)				(0.105)			
Retraction (r_t) x Ret Info	0.007	-0.409	0.025	-0.383	0.031	-0.405	0.025	-0.400
	(0.034)	(0.425)	(0.021)	(0.285)	(0.033)	(0.425)	(0.021)	(0.287)
Retracted Draw x Ret Info	-0.021	–	–	–	-0.006	–	–	–
	(0.128)				(0.128)			
Mean Decision Time			12.018				12.018	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.29	0.10	0.01	0.04	0.29	0.10	0.01	0.05
N	34137	34137	34137	4544	34137	34137	34137	4544

Clustered standard errors at the subject level in parentheses.

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

Table 23: Heterogenous Treatment Effects: Retraction Info

Notes: This table examines how treatment effects are affected by the provision of additional information about retractions. "Ret Info" corresponds to an indicator variable that equals 1 when the observation is from our treatment with additional retraction information in experiment C, and equals 0 when it is from our baseline treatment in experiment C. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline and "Retraction Info" treatments in experiment C, in which we collect such confidence level measure, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

Online Appendix F. Robustness 2: Consistency Across Heterogeneity

This section supports Section 6.2 of the paper. In Section F.1, we report regression tables corresponding to Figure 8. In Section F.2 we investigate heterogeneity in retraction ineffectiveness by subject, reporting summary statics on the subject-level estimates of the coefficient of interest (Retracted draw) in Table 2.

F.1. Regression Tables Corresponding to Figure 8

We include tables for each of the robustness checks presented in the figure, which explore heterogeneity in updating from retractions across multiple dimensions. For each dimension, we present two tables. The first re-estimates our main specification but with the sample restricted to the group in question (corresponding to the figure), while the second expands our main specification with interaction terms to account for heterogeneity, estimating the resulting regression on the full Baseline sample.

F.1.1. High Quantitative Ability

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.014	2.019***	0.097***	0.868***	-0.033	1.384***	0.109***	0.281
	(0.022)	(0.469)	(0.021)	(0.232)	(0.032)	(0.461)	(0.026)	(0.274)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.487***	–	–	–	0.563***	–	–	–
	(0.092)				(0.098)			
Mean Decision Time			8.299				8.299	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.40	0.13	0.02	0.02	0.41	0.14	0.02	0.02
N	11044	11044	11044	1470	11044	11044	11044	1470

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

Table 24: Treatment Effects: High Quant Ability

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects with above median score on a quantitative test in the experiment, corresponding to "High Quant Ability" of Figure 8. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.015	2.899***	0.051***	1.300***	-0.017	1.208***	0.072***	0.639***
	(0.023)	(0.316)	(0.014)	(0.205)	(0.028)	(0.343)	(0.015)	(0.205)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.628***	–	–	–	0.644***	–	–	–
	(0.086)				(0.103)			
Retraction (r_t) x High Quant	-0.015	-0.461	0.045*	-0.213	-0.008	-0.483	0.044*	-0.213
	(0.031)	(0.487)	(0.024)	(0.235)	(0.030)	(0.488)	(0.023)	(0.234)
Retracted Draw x High Quant	-0.149	–	–	–	-0.143	–	–	–
	(0.126)				(0.125)			
Mean Decision Time			8.830				8.830	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.27	0.12	0.01	0.03	0.27	0.12	0.01	0.04
N	39162	39162	39162	5236	39162	39162	39162	5236

Clustered standard errors at the subject level in parentheses.

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

Table 25: Heterogeneous Treatment Effects: Quant Ability

Notes: This table examines how treatment effects are affected by the subjects' quantitative ability. 'High Quant Ability' corresponds to an indicator variable that equals 1 when the observation is from subjects with above median score on a quantitative test in the experiment. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

F.1.2. High Confidence

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.010	2.560***	0.110***	1.257***	-0.094*	0.561	0.089***	0.537
	(0.050)	(0.447)	(0.022)	(0.343)	(0.054)	(0.528)	(0.029)	(0.358)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.554***	–	–	–	0.548***	–	–	–
	(0.126)				(0.185)			
Mean Decision Time			12.105				12.105	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.34	0.10	0.01	0.06	0.35	0.11	0.01	0.07
N	8130	8130	8130	1079	8130	8130	8130	1079

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

Table 26: Treatment Effects: High Confidence

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects with above median confidence in their beliefs, corresponding to "High Confidence" of Figure 8. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.057**	2.258***	0.047**	1.302***	0.014	0.717	0.029	0.611
	(0.027)	(0.441)	(0.023)	(0.361)	(0.038)	(0.451)	(0.022)	(0.378)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.644***	–	–	–	0.718***	–	–	–
	(0.131)				(0.147)			
Retraction (r_t) x Confident	-0.077	0.295	0.062**	-0.161	-0.074	0.264	0.066**	-0.180
	(0.053)	(0.580)	(0.029)	(0.428)	(0.051)	(0.582)	(0.029)	(0.425)
Retracted Draw x Confident	-0.085	–	–	–	-0.155	–	–	–
	(0.183)				(0.183)			
Mean Decision Time			11.765				11.765	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.28	0.11	0.01	0.05	0.29	0.11	0.01	0.06
N	16584	16584	16584	2206	16584	16584	16584	2206

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

Table 27: Heterogeneous Treatment Effects: Confidence

Notes: This table examines how treatment effects are affected by the subjects' confidence. "High Quant Ability" corresponds to an indicator variable that equals 1 when the observation is from subjects with above median confidence in their beliefs. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline treatment in experiment C, for which we collect confidence data, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

F.1.3. More Bayesian

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.014	2.828***	0.075***	0.794***	0.003	1.835***	0.109***	0.465***
	(0.015)	(0.318)	(0.017)	(0.148)	(0.020)	(0.296)	(0.019)	(0.147)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.571***	–	–	–	0.651***	–	–	–
	(0.057)				(0.067)			
Mean Decision Time			8.736				8.736	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.52	0.16	0.01	0.05	0.53	0.17	0.02	0.06
N	20972	20972	20972	2791	20972	20972	20972	2791

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

Table 28: Treatment Effects: More Bayesian

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to subjects who have, on average, lower than median distance to the Bayesian posterior when updating from new draws, corresponding to "More Bayesian" of Figure 8. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.003	2.285***	0.046***	1.481***	-0.028	0.542	0.065***	0.819***
	(0.034)	(0.404)	(0.016)	(0.276)	(0.038)	(0.400)	(0.018)	(0.274)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.611***	–	–	–	0.618***	–	–	–
	(0.128)				(0.141)			
Retraction (r_t) x More Bayesian	0.026	0.877*	0.034	-0.461*	0.017	0.925*	0.036*	-0.458
	(0.037)	(0.485)	(0.022)	(0.279)	(0.036)	(0.484)	(0.021)	(0.279)
Retracted Draw x More Bayesian	-0.045	–	–	–	-0.025	–	–	–
	(0.140)				(0.139)			
Mean Decision Time			8.830				8.830	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.27	0.27	0.01	0.05	0.27	0.27	0.01	0.06
N	39162	39162	39162	5236	39162	39162	39162	5236

Clustered standard errors at the subject level in parentheses.

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

Table 29: Heterogenous Treatment Effects: Bayesian

Notes: This table examines how treatment effects are affected by how correctly subjects update beliefs from new draws. "More Bayesian" corresponds to an indicator variable that equals 1 when the observation is from subjects more Bayesian than the median subject when updating from new draws (i.e., have on average lower than median distance to the Bayesian posterior). There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

F.1.4. More Experienced (2nd Half)

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.039*	2.743***	0.052***	0.997***	-0.038	1.206***	0.059***	0.712***
	(0.020)	(0.309)	(0.013)	(0.258)	(0.029)	(0.352)	(0.016)	(0.233)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.585***	–	–	–	0.550***	–	–	–
	(0.080)				(0.106)			
Mean Decision Time			7.414				7.414	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.28	0.07	0.01	0.02	0.28	0.07	0.01	0.02
N	20834	20834	20834	4054	20834	20834	20834	4054

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

Table 30: Treatment Effects: Experienced

Notes: This table reports the effect of retractions on updating and empirical complexity measures when restricting the baseline sample to the second half of rounds for each subject, corresponding to "Experienced" of Figure 8. Columns (1) and (5) are the regressions from Table 2, but restricted to this sample. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	-0.007	2.675***	0.080***	0.817***	-0.038	1.020***	0.098***	0.541***
	(0.026)	(0.287)	(0.016)	(0.144)	(0.031)	(0.302)	(0.017)	(0.136)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.587***	–	–	–	0.608***	–	–	–
	(0.079)				(0.094)			
Retraction (r_t) x Experienced	0.034	0.169	-0.032*	0.235	0.037	0.163	-0.032**	0.240
	(0.029)	(0.272)	(0.016)	(0.236)	(0.029)	(0.270)	(0.016)	(0.232)
Retracted Draw x Experienced	-0.001	–	–	–	-0.008	–	–	–
	(0.086)				(0.084)			
Mean Decision Time			8.830				8.830	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.26	0.08	0.03	0.02	0.27	0.08	0.04	0.02
N	39162	39162	39162	7822	39162	39162	39162	7822

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

Table 31: Heterogenous Treatment Effects: Experience

Notes: This table examines how treatment effects are affected by experience. "Experienced" corresponds to an indicator variable that equals 1 when the observation is from the second half of the rounds. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

F.2. Individual Heterogeneity

Retraction vs.	No Retracted Draw	Equivalent New Draw
	(1)	(2)
	$\hat{\ell}_t$	$\hat{\ell}_t$
Mean Subject-level Effect	0.628	0.564
	(0.080)	(0.106)
Median Subject-level Effect	0.340	0.297
	(0.059)	(0.058)
Fraction $\beta_1 > 0$	0.720	0.683
Mean Std Error	0.462	0.502
Median Std Error	0.268	0.274
Compressed History FEs	Yes	No
Sign History FEs	No	Yes

Bootstrapped standard errors in parentheses.

Table 32: Updating from Retractions (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 2 for each subject. The sample includes all observations of subjects in the baseline treatment, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

Online Appendix G. Robustness 3: Variations on the Design

This section supports Section 6.3 of the paper. In Section G.1, we report regression tables corresponding to Figure 9.

G.1. Regression Tables Corresponding to Figure 9

We include tables for each of the robustness checks presented in the figure, which explore the robustness of our results to variations in the experimental design. For each variation, we re-estimate our main specification using the results from the variation of the design (corresponding to the figure). The first and second variants were randomized against the baseline variant, at the individual subject level. Hence, for those variants, we also present results on heterogeneous treatment effects, directly using an interaction term.

G.1.1. Elicit at End

Retraction vs.	Equivalent New Draw			
	(1)	(2)	(3)	(4)
	$\hat{\ell}_t$	$\lvert\hat{p}_t - p_t\rvert$	$\ln(T_t)$	$\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.034	-0.916**	0.125***	0.299**
	(0.041)	(0.456)	(0.028)	(0.117)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.615***	–	–	–
	(0.140)			
Mean Decision Time			9.769	
Sign History FEs	Yes	Yes	Yes	Yes
R^2	0.42	0.03	0.01	0.02
N	6093	6093	6093	1436

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

Table 33: Treatment Effects: Elicit at End

Notes: This table reports the effect of retractions on updating and empirical complexity measures in a variant of the experiment in which beliefs are elicited only at the end of each round, corresponding to "Elicit at End" of Figure 9. Column (1) is the equivalent regression from Table 2. Similarly, Columns (2)-(4) are the regressions from Table 3. We cannot compare to 'No Retracted Draw' as beliefs are only elicited in the last period of each round.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\text{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.008	3.050***	0.053***	1.176***	-0.016	0.825**	0.109***	0.593***
	(0.023)	(0.391)	(0.016)	(0.240)	(0.032)	(0.406)	(0.015)	(0.201)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.572***	–	–	–	0.584***	–	–	–
	(0.095)				(0.113)			
Retraction (r_t) x Elicit End	0.040	-1.053*	-0.054*	-0.317	0.046	-0.650	-0.006	-0.243
	(0.051)	(0.552)	(0.028)	(0.204)	(0.051)	(0.538)	(0.028)	(0.199)
Retracted Draw x Elicit End	-0.071	–	–	–	0.062	–	–	–
	(0.152)				(0.149)			
Mean Decision Time			7.332				7.332	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.29	0.05	0.07	0.02	0.29	0.05	0.07	0.03
N	28671	28671	28671	4466	28671	28671	28671	4466

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

Table 34: Heterogenous Treatment Effects: Elicit at End

Notes: This table examines how treatment effects are affected by eliciting beliefs only at the end of the round. "Elicit at End" corresponds to an indicator variable that equals 1 when the observation is from that treatment in experiment A, and equals 0 when it is from the baseline treatment in experiment A. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline and 'Elicit at the End' treatments of experiment A, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

G.1.2. No History

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.018	5.043***	0.059***	1.703***	0.029	3.915***	0.088***	0.875*
	(0.059)	(0.572)	(0.017)	(0.429)	(0.056)	(0.558)	(0.020)	(0.461)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.392**	–	–	–	1.034***	–	–	–
	(0.195)				(0.273)			
Mean Decision Time			12.000				12.000	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.16	0.04	0.01	0.02	0.16	0.04	0.01	0.02
N	17642	17642	17642	2342	17642	17642	17642	2342

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

Table 35: Treatment Effects: No History

Notes: This table reports the effect of retractions on updating and empirical complexity measures in a variant of the experiment in which subjects were only shown the current observation, not the history of all observations in the current round, corresponding to "No History" of Figure 9. Columns (1) and (5) are the equivalent regressions from Table 2. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(\mathrm{T}_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(\mathrm{T}_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.005	2.253***	0.070***	1.028***	-0.012	0.907***	0.078***	0.258
	(0.027)	(0.302)	(0.016)	(0.227)	(0.031)	(0.322)	(0.017)	(0.258)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.601***	–	–	–	0.954***	–	–	–
	(0.090)				(0.124)			
Retraction (r_t) x No Hist	0.026	2.918***	-0.004	0.853*	0.032	2.918***	-0.005	0.853*
	(0.062)	(0.622)	(0.021)	(0.453)	(0.062)	(0.621)	(0.021)	(0.452)
Retracted Draw x No Hist	-0.209	–	–	–	-0.203	–	–	–
	(0.215)				(0.215)			
Mean Decision Time			11.886				11.886	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.21	0.07	0.01	0.02	0.21	0.07	0.01	0.02
N	34226	34226	34226	4548	34226	34226	34226	4548

Clustered standard errors at the subject level in parentheses.

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

Table 36: Heterogenous Treatment Effects: No History

Notes: This table examines how treatment effects are affected by not showing the history of past signals. "No History" corresponds to an indicator variable that equals 1 when the observation is from that treatment in experiment C, and equals 0 when it is from the baseline treatment in experiment C. There are two types of comparison: (a) updating from a retraction vs. without the retracted observation (Columns (1)-(4)) and (b) vs. an equivalent new draw (Columns (5)-(8)). Columns (1) and (5) show effects on log-odds beliefs; (2) and (6) on the accuracy of belief updating; (3) and (7) on the speed of updating; (4) and (8) on the variability of updating. The sample includes all observations of subjects in the baseline and 'No History' treatments of experiment C, excluding periods in which the truth ball is disclosed or in which there was a retraction in an earlier period.

G.1.3. Short Histories

Retraction vs.	No Retracted Draw		Equivalent New Draw			
	(1)	(2)	(3)	(4)	(5)	(6)
	$\hat{\ell}_t$	$\lvert\hat{p}_t - p_t\rvert$	$\hat{\ell}_t$	$\lvert\hat{p}_t - p_t\rvert$	$\ln(T_t)$	$\text{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.001	1.525**	-0.013	6.378***	0.081**	1.123***
	(0.025)	(0.697)	(0.027)	(0.796)	(0.032)	(0.434)
Retracted Draw ($r_t \cdot K(s_{\rho_t})$)	0.448***	–	0.371**	–	–	–
	(0.169)		(0.166)			
Mean Decision Time					8.131	
Compressed History FEs	Yes	Yes	No	No	No	No
Sign History FEs	No	No	Yes	Yes	Yes	Yes
R^2	0.01	0.00	0.21	0.08	0.00	0.02
N	13938	13938	9138	9138	9138	882

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

Table 37: Treatment Effects: Short Histories

Notes: This table reports the effect of retractions on updating and empirical complexity measures in a variant of the experiment in which there were only two periods per round, rather than four, corresponding to "Short Histories" of Figure 9. Columns (1) and (3) are the equivalent regressions from Table 2. Similarly, Columns (2) and (4)-(6) are the regressions from Table 3. There are no treatment effects on decision times and variability for the comparison to 'No Retracted Draw', because, given there are only two periods, the comparison is to the prior at period 0 (i.e. before any observations) which we did not elicit and assume here to be 0.5.

G.1.4. Garbled Information

Retraction vs.	No Retracted Draw				Equivalent New Draw			
	(1) $\hat{\ell}_t$	(2) $\lvert\hat{p}_t - p_t\rvert$	(3) $\ln(T_t)$	(4) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$	(5) $\hat{\ell}_t$	(6) $\lvert\hat{p}_t - p_t\rvert$	(7) $\ln(T_t)$	(8) $\mathrm{Var}(\hat{\ell}_t \mid h_t)$
Retraction (r_t)	0.001	3.164***	0.090***	0.660**	-0.007	3.591***	0.116***	0.242
	(0.026)	(0.610)	(0.027)	(0.277)	(0.040)	(0.647)	(0.028)	(0.307)
Retracted Draw ($r_t \cdot K(s_{\rho t})$)	0.338***	–	–	–	0.567***	–	–	–
	(0.067)				(0.086)			
Mean Decision Time			8.761				8.761	
Compressed History FEs	Yes	Yes	Yes	Yes	No	No	No	No
Sign History FEs	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.33	0.04	0.01	0.01	0.33	0.04	0.01	0.01
N	14427	14427	14344	1703	14427	14427	14344	1703

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

Table 38: Treatment Effects: Garbled Information

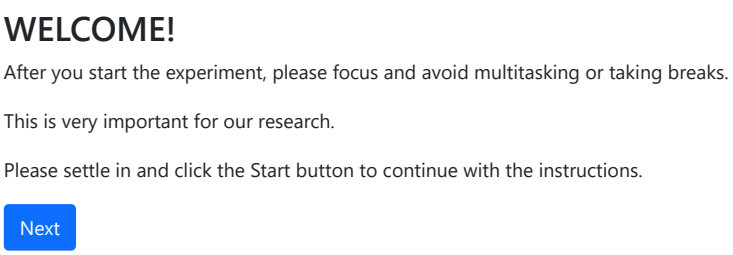
Notes: This table reports the effect of retractions on updating and empirical complexity measures in a variant of the experiment in which truth balls were not fully informative, corresponding to "Garbled Info" of Figure 9. Columns (1) and (5) are the equivalent regressions from Table 2. Similarly, Columns (2)-(4) and (6)-(8) are the regressions from Table 3.

Online Appendix H. Instructions and Screenshots

All our treatments followed small variations of our baseline instructions presented here.

H.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:

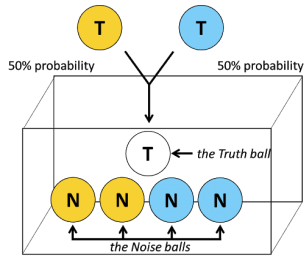
- 4 Noise Balls (N), of which 2 are yellow (N) and 2 are blue (N); and
- 1 Truth Ball (T), which can be either yellow (T) or blue (T).

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.



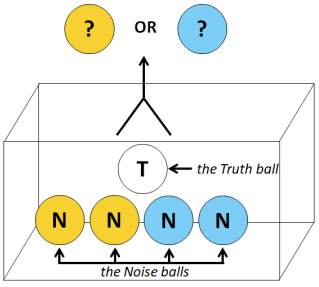
The round lasts for 4 periods, each of which may help you to guess the color of the Truth Ball (T).

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

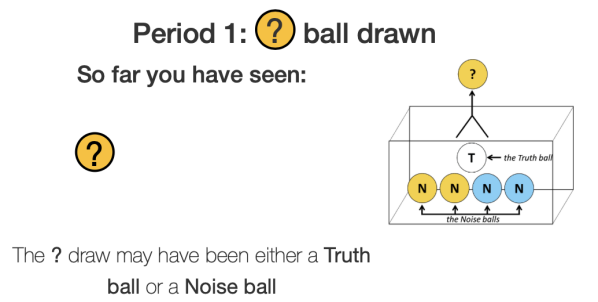
In periods 1 and 2, a ball is drawn from the box at random and you are told its color, (●) or (●).
The ball is then placed back into the box.
You will not be told whether it is a Noise Ball (N) or the Truth Ball (T). Because of this, the ball will be labelled with a question mark (?).
Since the balls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

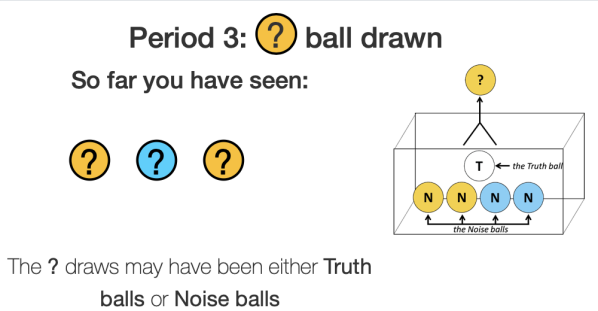
(i) New Draw

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:



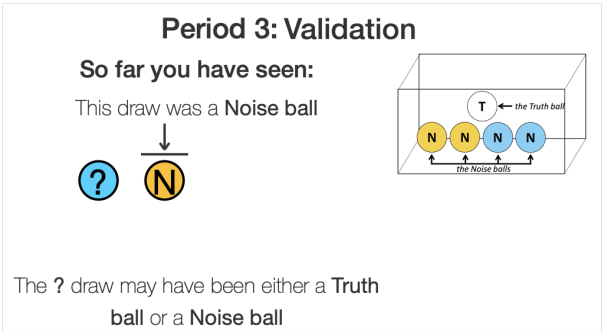
(ii) Validation

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 Ⓣ is selected the same way and independently.

This means that whether the Truth Ball is Ⓣ or Ⓣ in one round has no influence on whether the Truth Ball is Ⓣ or Ⓣ in another round.

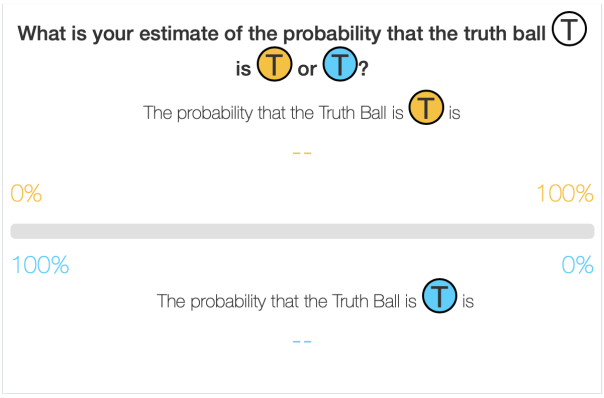
It will be clearly indicated when a new round begins.

Estimates

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

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 Ⓣ 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  and  ?

○ 1  and 3 

○ 3  and 1 

○ 2  and 2 

4. It is possible that you see a  ball 3 times and the Truth Ball is  .

○ The statement is true.

○ The statement is false.

5. Even if in a given round the Truth Ball is  , in the following round the Truth Ball can be either  or  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 then the Truth Ball has to be  and not  .

○ The statement is true.

○ The statement is false.

7. If a draw you were shown  corresponded to a Noise Ball  then the Truth Ball  may or may not be of a different color.

○ The statement is true.

○ The statement is false.

Check Answers

H.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

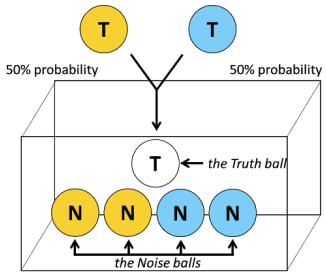
You will now play two practice rounds.
These rounds do not count towards your payment.
They are meant for you to familiarize yourself with the interface and the task.

Start Practice Rounds

Practice Round 1 of 2

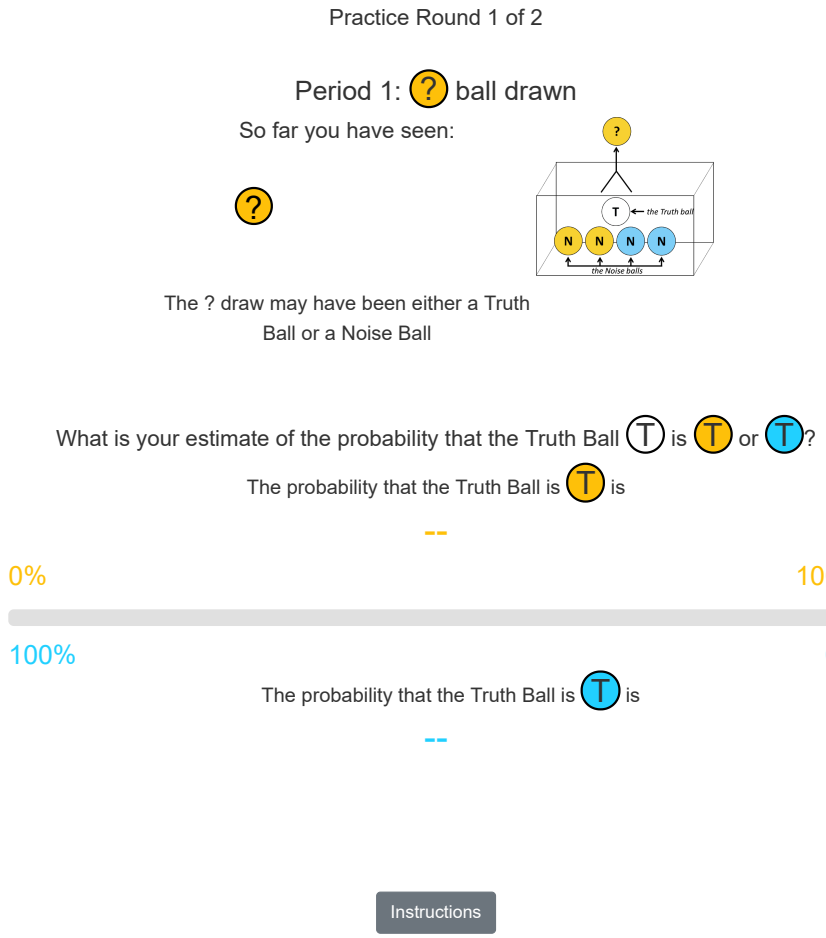
New Round

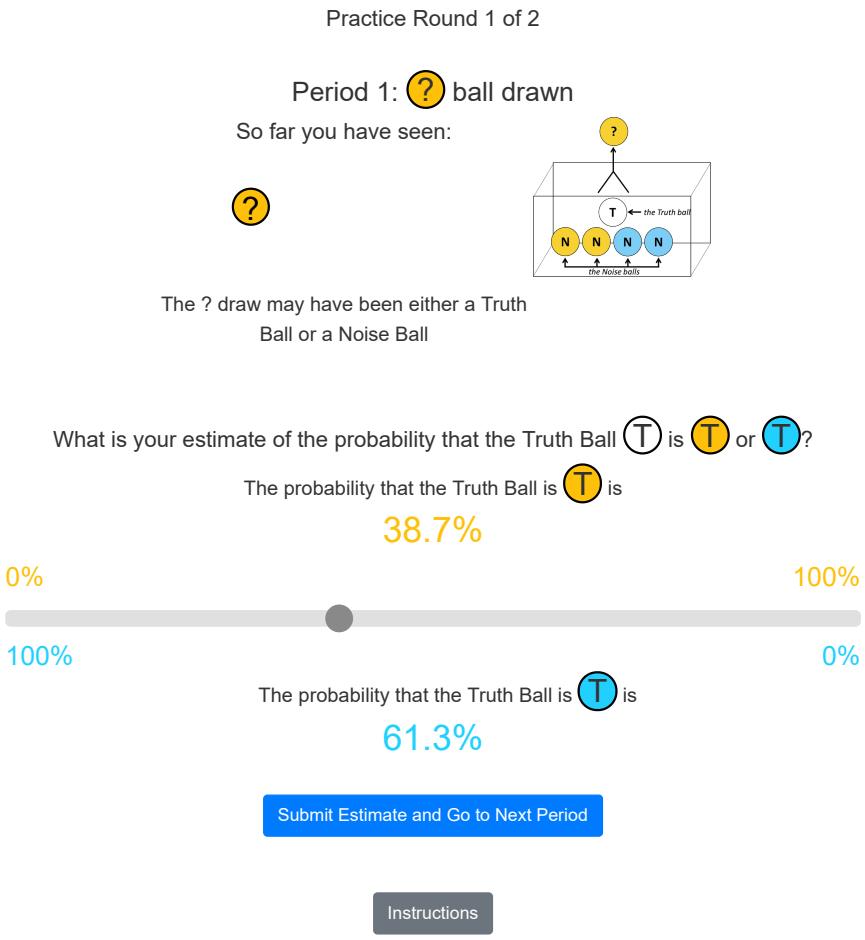
The truth ball is drawn and placed in the box



Start New Round

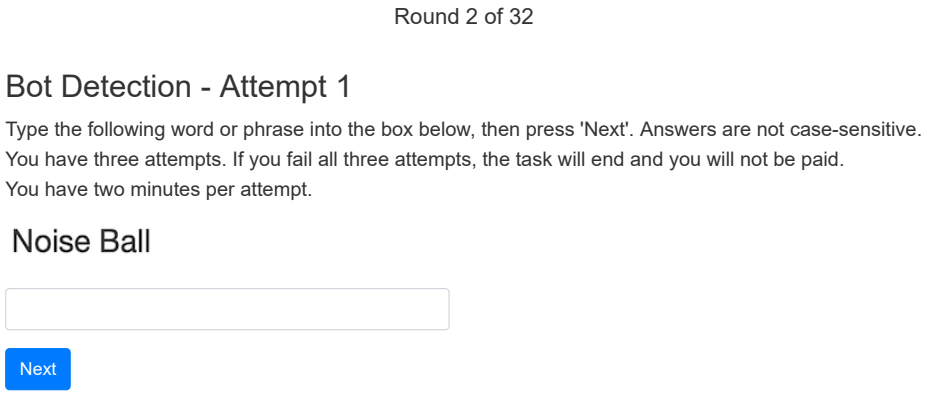
One the page loaded, the slider was blank and only activated once the subjects clicked on it.





H.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.



H.4. Rounds

The rounds were described in Section 2.2.

Start the Task

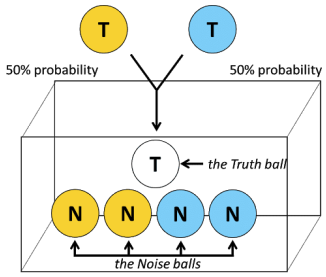
From now on, rounds matter towards your payment.

Start the Task

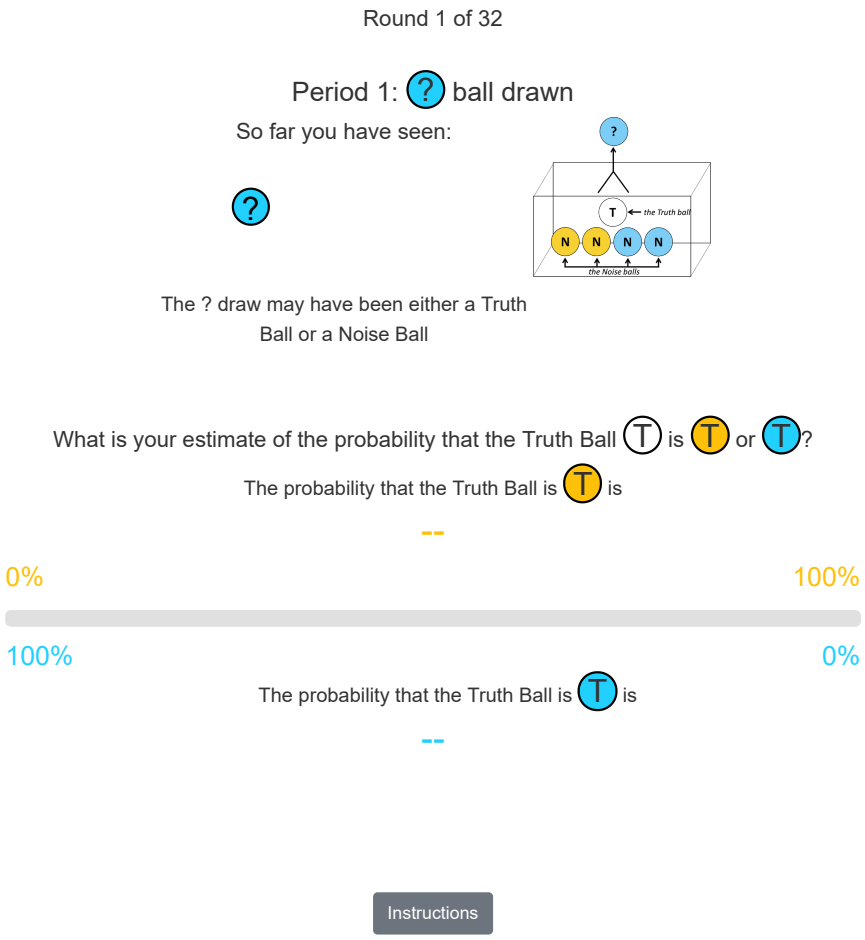
Round 1 of 32

New Round

The truth ball is drawn and placed in the box

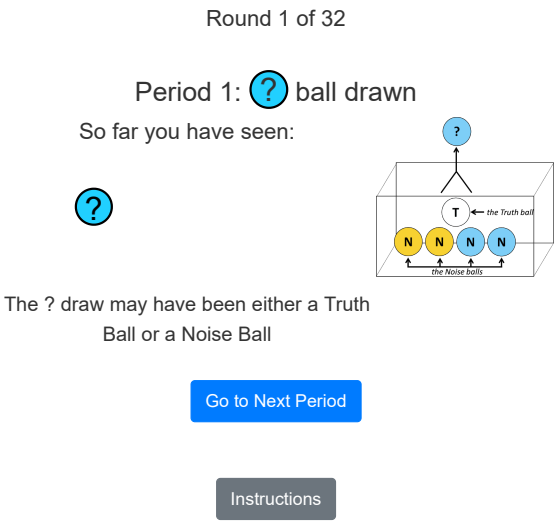


Start New Round



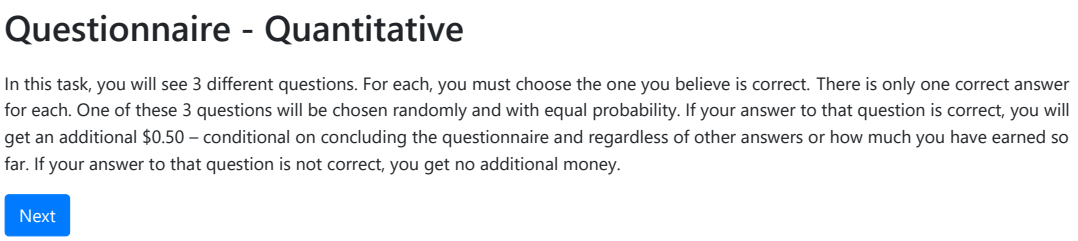
H.5. Final Period Elicitation Only

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:



H.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

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.

H.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.