Belief Updating and Misinformation

ESA - Santa Barbara

Lars Wittrock (joint with Elias Tsakas and Martin Strobel)

November 12, 2022

Maastricht University

Introduction

Two general observations

- New information is sometimes **not fully reliable at first**.
  - News reports (PewResearch, 2022).
  - Factual claims in discussions (e.g. politicians or friends/family).
  - Information leaks from anonymous sources.
  - Academic research on new topics (e.g. Covid-19).

Two general observations

- New information is sometimes **not fully reliable at first**.
  - News reports (PewResearch, 2022).
  - Factual claims in discussions (e.g. politicians or friends/family).
  - Information leaks from anonymous sources.
  - Academic research on new topics (e.g. Covid-19).
- Uncertain information is frequently **confirmed or retracted later**.

Two general observations

The New York Times

USA Today to Remove 23 Articles After Investigation Into Fabricated Sources

The articles were removed after an investigation identified stories with sources that appeared to be fabricated, USA Today said.

- New ir
  - N
  - Fa … amily).
  - In
  - A

- Uncertain information is frequently **confirmed or retracted later**.

Two general observations

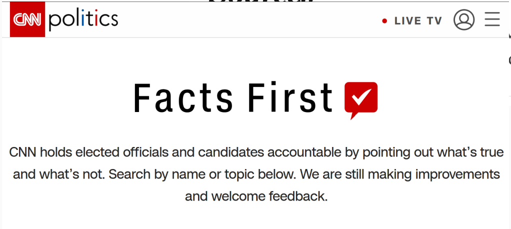
- New ir
  - N

amily).


The New York Times

USA Today to Remove 23 Articles After Investigation Into Fabricated Sources

vestigation identified stories
cated, USA Today said.


CNN politics • LIVE TV

Facts First

CNN holds elected officials and candidates accountable by pointing out what's true and what's not. Search by name or topic below. We are still making improvements and welcome feedback.

confirmed or retracted later.

Two general observations

The New York Times

USA Today to Remove 23 Articles After Investigation Into Fabricated Sources

- New ir
  - N

…vestigation identified stories …cated, USA Today said.

…amily).

CNN politics

LIVE TV

Facts First

CNN holds elected officials and candidates accountal… and what's not. Search by name or topic below. We a… and welcome feedback…

THE LANCET

Submit Article Log in Register Subscribe

RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Prof Mandeep R Mehra, MD • Sapan S Desai, MD • Prof Frank Ruschitzka, MD • Amit N Patel, MD

Published: May 22, 2020 • DOI: https://doi.org/10.1016/S0140-6736(20)31180-6 • Check for updates

Summary

References

Article Info

Figures

Tables

Linked Articles

Related Specialty

Summary

Background

RETRACTED

Hydroxychloroquine or chloroquine, often in combination with a second-generation macrolide, are being widely used for treatment of COVID-19, despite no conclusive evidence of their benefit. Although generally safe when used for approved indications such as autoimmune disease or malaria, the safety and benefit of these treatment regimens are poorly evaluated in COVID-19.

Two general observations

The New York Times

USA Today to Remove 23 Articles After Investigation Into Fabricated Sources

- New ir
  - N

CNN politics

Facts First

CNN holds elected officials and candidates accountal
and what's not. Search by name or topic below. We a
and welcome feedback

THE

Twitter confirms data from 5.4 million accounts has been stolen

Twitter has confirmed that a suspected data breach in July led to account data being stolen

RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Prof Mandeep R Mehra, MD • Sapan S Desai, MD • Prof Frank Ruschitzka, MD • Amit N Patel, MD

Published: May 22, 2020 • DOI: https://doi.org/10.1016/S0140-6736(20)31180-6 • Check for updates

Summary
References
Article Info
Figures
Tables
Linked Articles
Related Specialty

Summary

Background

Hydroxychloroquine or chloroquine, often in combination with a second-generation macrolide, are being widely used for treatment of COVID-19, despite no conclusive evidence of their benefit. Although generally safe when used for approved indications such as autoimmune disease or malaria, the safety and benefit of these treatment regimens are poorly evaluated in COVID-19.

RETRACTED

Motivation

- Reaction to new information is well studied (Benjamin, 2019).
- It is unclear how people deal with *information about information*, i.e. confirmations or retractions.
    - Significant differences between people in their acceptance of misinformation after retractions (Meyer et al., 2020).

Research Questions

1. How do people update their belief when being told a previous signal was fully uninformative?
    - Continued Influence Effect in psychology and Goncalves et al. (2022) show (small) average effect.
    - Mechanism not clear.

2. How do people update their belief when being told a previous signal was indeed informative? [Not part of today]
    - No prior evidence.

Literature

Framework

What do we want from our design?

Requirements:

- Introduce information uncertainty.
- Neutral setting without motivated beliefs.
- Verifications of previous information are unambiguous.
- Belief elicitation can be incentivized.
- Bayesian beliefs can be computed.
- (Results can be compared to the literature).

What do we want from our design?

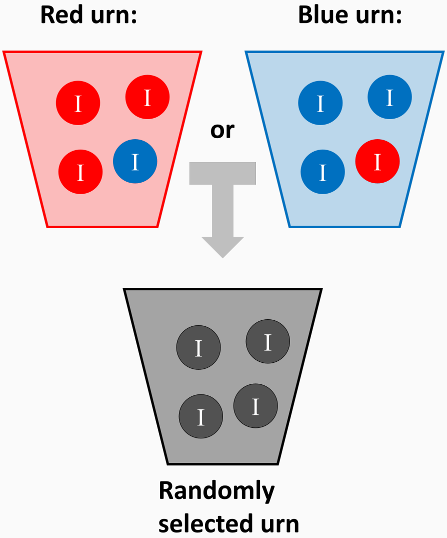
Requirements:

- Introduce information uncertainty.
- Neutral setting without motivated beliefs.
- Verifications of previous information are unambiguous.
- Belief elicitation can be incentivized.
- Bayesian beliefs can be computed.
- (Results can be compared to the literature).

$\Rightarrow$ Modified ball and urn framework

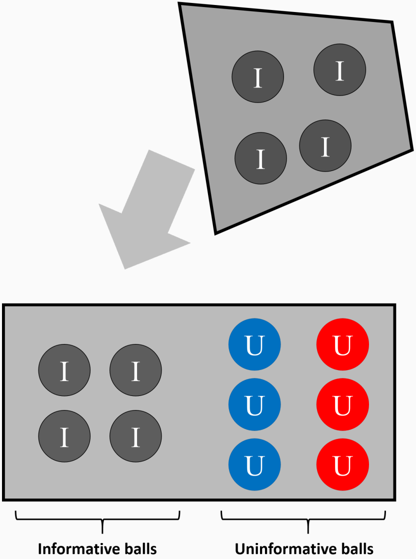
Modified ball and urn framework

Step 1:



Modified ball and urn framework

Step 2:



Hints

Two types of hints:

- *Regular:* Color of ball shown. Example: 

- *Check:* Told if the previous ball was 'informative' (I) or 'uninformative' (U). The previous ball is again displayed.

Number of hints:

- 9 regular signals and 3 verifications.
  - Verifications are always immediately after the respective ball.
- Which balls are verified varies per subject.

Example Screen

Round 7

Background:

Show/hide instructions

History:

Ball 1	Ball 2	Ball 3	Ball 4	Ball 5	Ball 6	Ball 7	Ball 8	Ball 9
?	?	?	?	U	?			

You previously thought it was **50%** likely that the selected urn is red.

New Information:

A **blue** ball was drawn from the black box: ? It is put back into the box with the other balls.

Question:

What do you think are the chances (in %) that the RED URN was picked in the beginning?

\- 0 100 +

%

Results

Sample Overview

Sample:

- 606 subjects completed the experiment on Prolific.
- 46 were removed as outliers (pre-registered criteria).
- In total 6,720 observations.
- Median time to complete survey 17 minutes.
- Average payoff is 4.80€.

Sanity check:

- Beliefs and Bayesian posteriors are highly correlated ($R^2 = 0.51$).

Regression More

How do people react to retractions?

Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

How do people react to retractions?

Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

Question: Influenced by initial update?

How do people react to retractions?

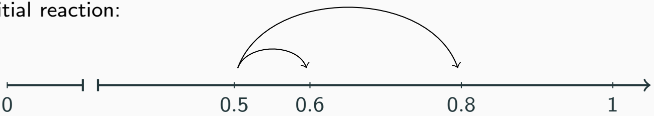
Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

Question: Influenced by initial update?

Example: retracted red ball

Initial reaction:



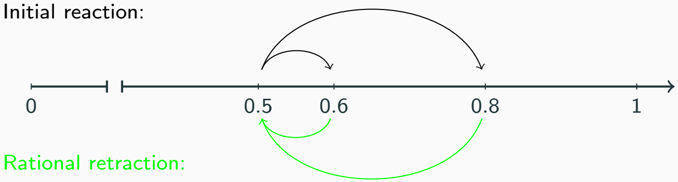
How do people react to retractions?

Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

Question: Influenced by initial update?

Example: retracted red ball



How do people react to retractions?

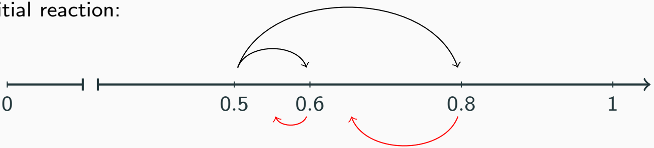
Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

Question: Influenced by initial update?

Example: retracted red ball

Initial reaction:



Continued influence:

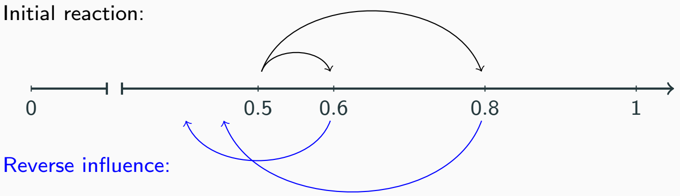
How do people react to retractions?

Rational Posterior

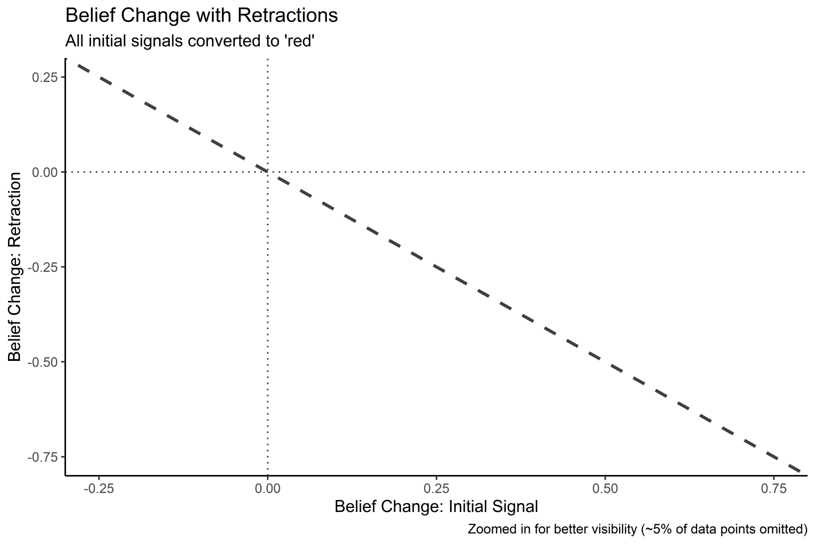
- Simply 'forget' the initial uncertain signal.
- Return to the prior belief before retracted signal.

Question: Influenced by initial update?

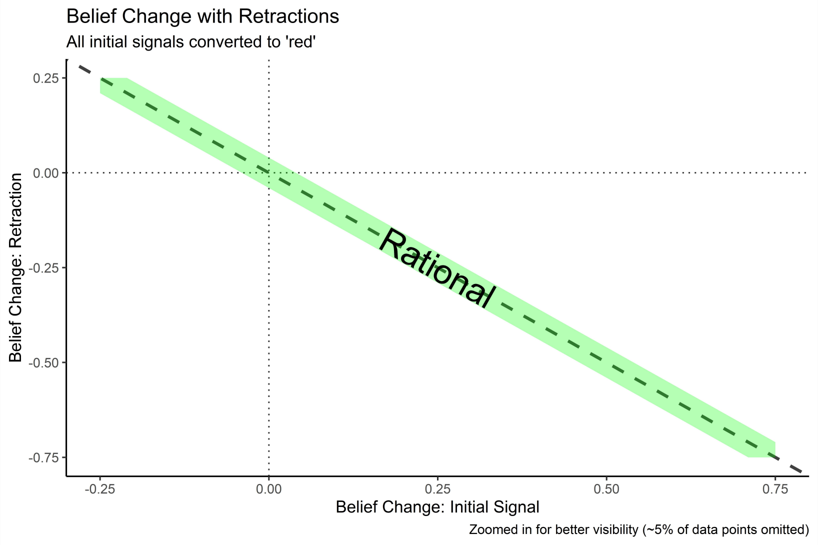
Example: retracted red ball



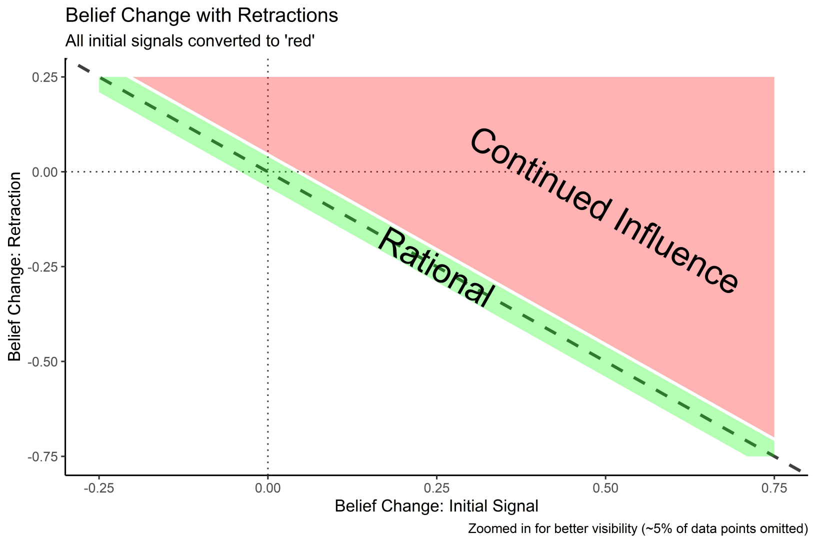
Initial Reaction vs Retraction



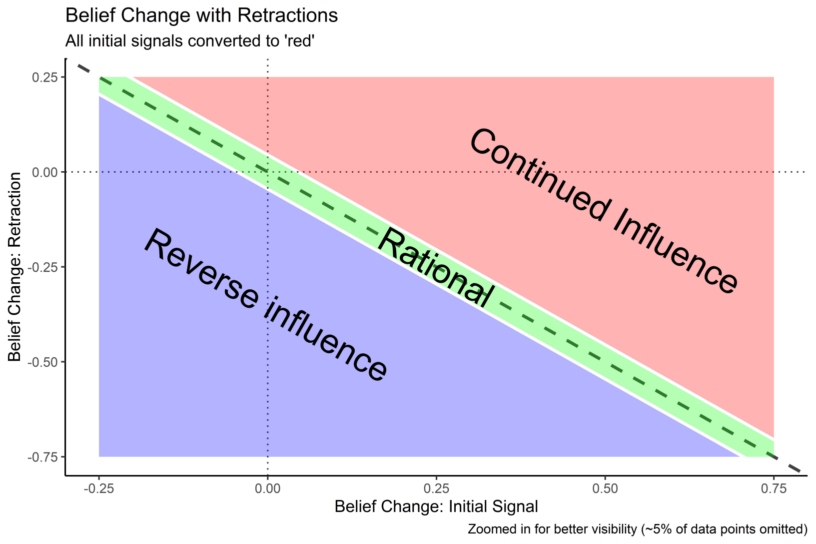
Initial Reaction vs Retraction



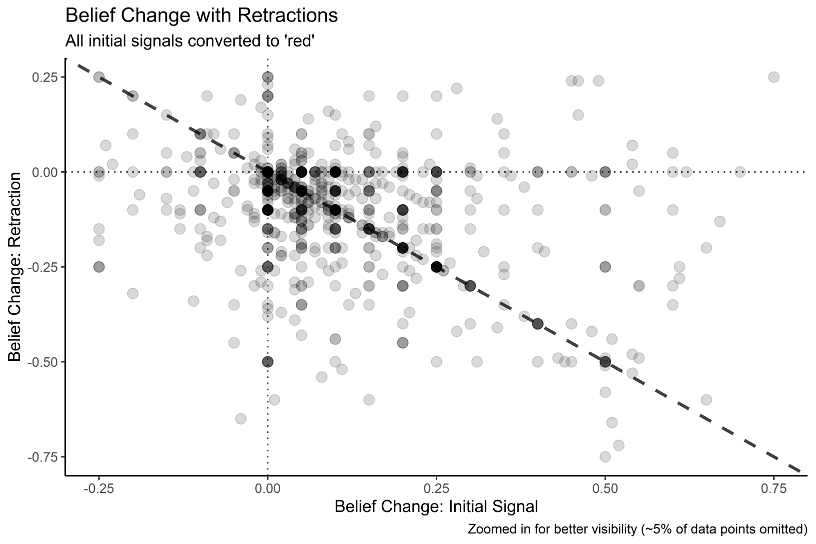
Initial Reaction vs Retraction



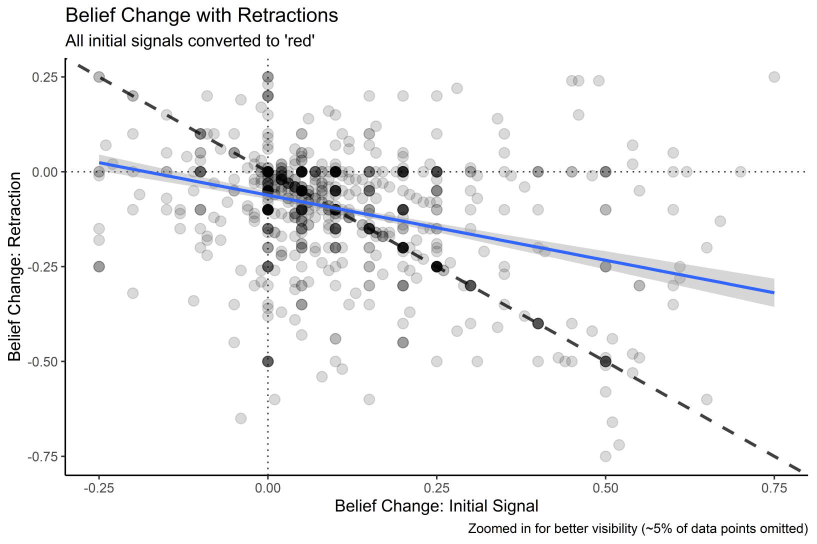
Initial Reaction vs Retraction



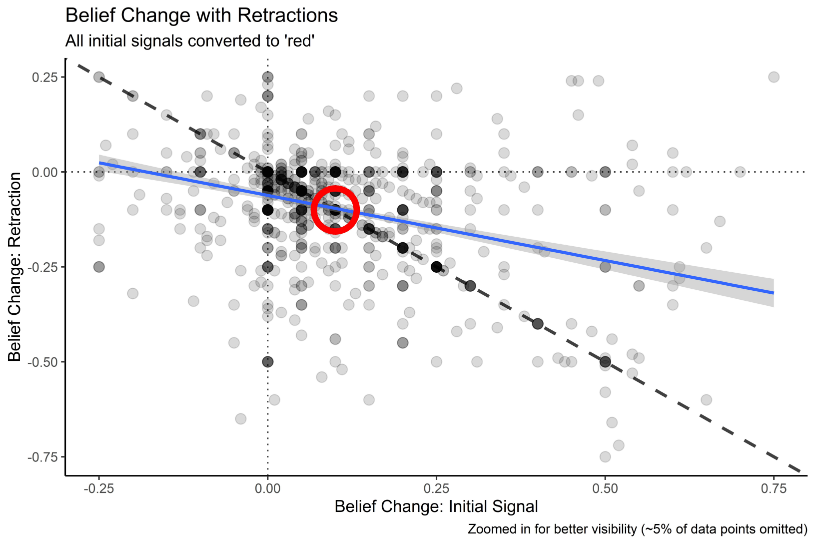
Initial Reaction vs Retraction



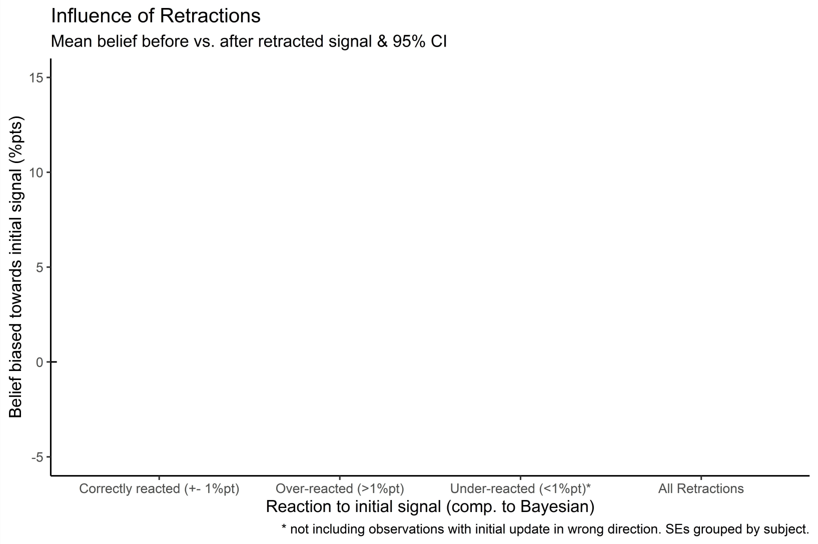
Initial Reaction vs Retraction



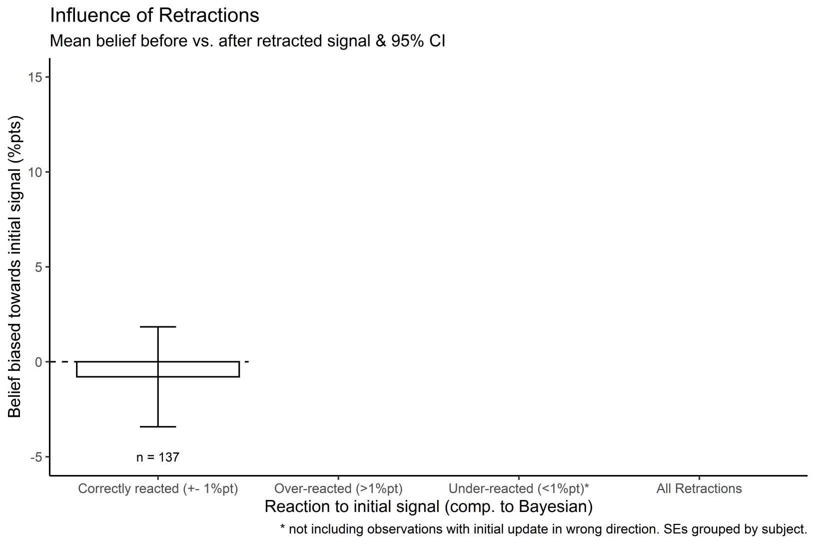
Initial Reaction vs Retraction



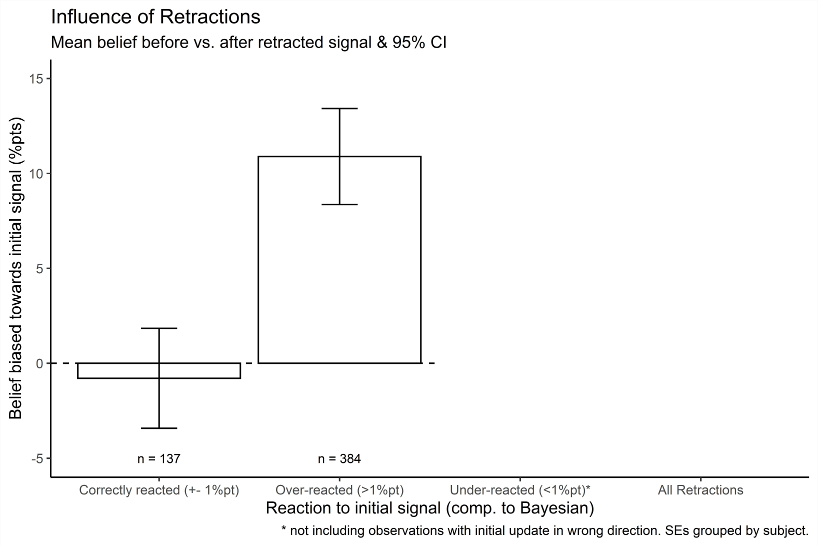
Initial update explains reaction to retractions



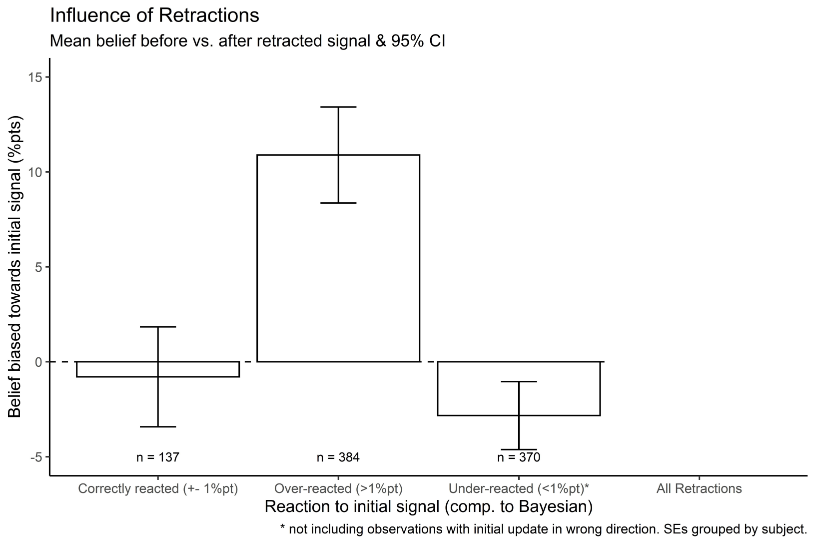
Initial update explains reaction to retractions



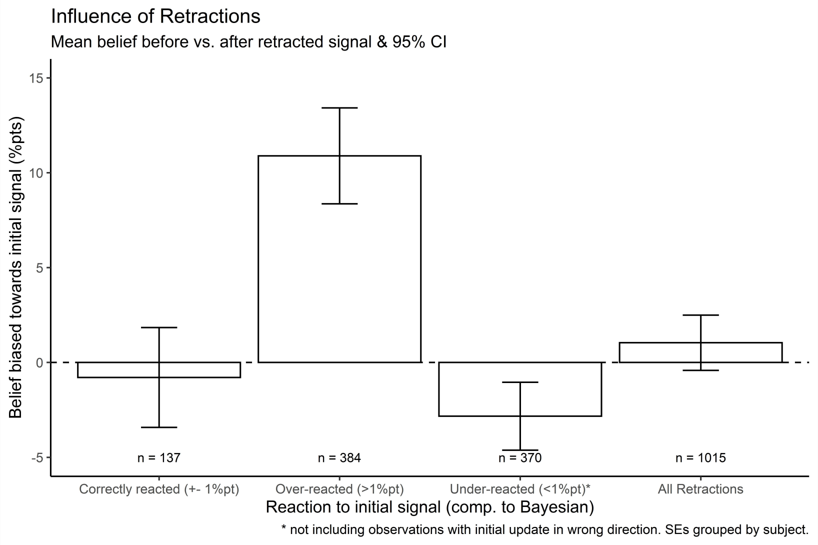
Initial update explains reaction to retractions



Initial update explains reaction to retractions

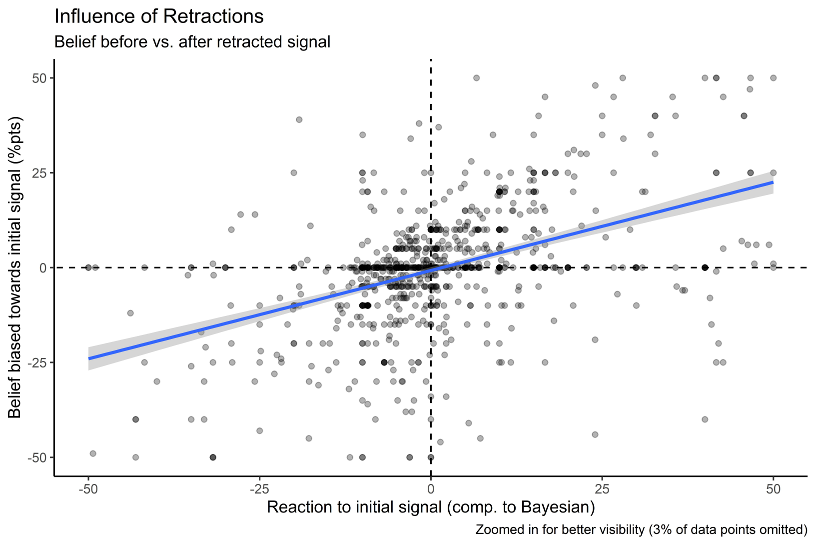


Initial update explains reaction to retractions



Magnitude of initial mistake matters

Regression



Alternative Mechanisms

Robustness checks:

- Are retractions different to 'regular' signals? ✓

Alternative Mechanisms

Robustness checks:

- Are retractions different to 'regular' signals? ✓
- Are subjects consistent? ✗

Alternative Mechanisms

Robustness checks:

- Are retractions different to 'regular' signals? ✓
- Are subjects consistent? ✗
- Anchoring? ✗

Alternative Mechanisms

Robustness checks:

- Are retractions different to 'regular' signals? ✓
- Are subjects consistent? ✗
- Anchoring? ✗
- Correction of previous mistake? ✗

Summary

Summary

Findings:

- Initial reaction to uncertain signals determines how people respond to their retraction (even in a neutral setting!):
  - Overly trusting initially: continued influence of retracted information.
  - Overly sceptical initially: reverse effect.

Implication:

- Misinformation (even if corrected immediately) is a potential reason for persisting polarized beliefs.
- Correcting information ex-post is only (fully) effective if people reacted correctly initially.
- Motivated beliefs are likely to amplify this effect.

Thank you for your attention!

Questions?

References

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics - Foundations and Applications 2*, pages 69–186. Elsevier.

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

Goncalves, D., Libgober, J., and Willis, J. (2022). Learning versus unlearning: An experiment on retractions. *NBER Working Paper*.

Meyer, M., Alfano, M., and de Bruin, B. (2020). Epistemic vice predicts acceptance of covid-19 misinformation. *SSRN Electronic Journal*.

PewResearch (2022). Journalists sense turmoil in their industry amid continued passion for their work.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3):346–354.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Appendix

Literature

- Belief updating problems first studied in 60s and 70s.
  - Phillips and Edwards (1966); Tversky and Kahneman (1971, 1974); and many others.
- Benjamin (2019) – Meta study
  - Strong evidence of under-inference and base-rate neglect.
  - Does not mention information uncertainty.
- Psychology literature - Continued Influence Effect[1]
  - People fail to 'unlearn' retracted information.
  - Framework: narratives with a causal structure.
  - Cognitive ability partly explains the size of the effect.
- Goncalves et al. (2022)
  - Subjects fail to 'unlearn' from retractions even in a neutral setting.
  - Mechanism: Retractions are harder to process than regular signals.

Back

[1] For an overview of the literature see Ecker et al. (2022).

Different Types of Information Signals

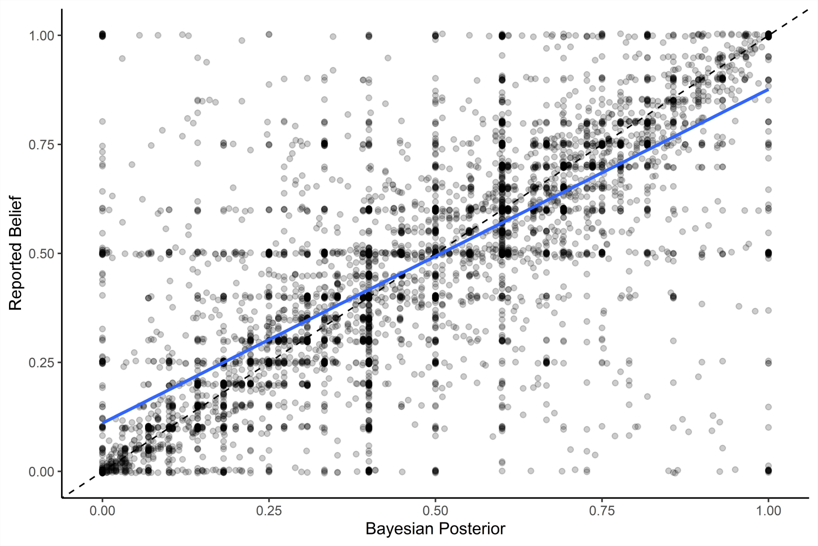
- Two states of the world: $\Theta = \{\text{Blue } (B),\ \text{Red } (R)\}$
- Two possible signal realizations: $S = \{\text{blue } (b),\ \text{red } (r)\}$

- Two signals:
  - Informative signal: $\pi_I(b|B) = \pi_I(r|R) = 1 - \varepsilon$, with $\varepsilon \leq 0.5$.
  - Noisy signal: $\pi_N(b|B) = \pi_N(b|R) = \beta$, for some $\beta \in [0, 1]$.

- Combining both signals: $\pi(b|\theta) = \alpha\pi_I(b|\theta) + (1 - \alpha)\beta$
- For the experiment we set $\alpha = 0.4$ and $\beta = 0.5$. Hence: $\pi(b|B) = \pi(r|R) = 0.6$

Back

Sanity Check: Reported Beliefs vs Bayesian Posteriors



Back

Sanity Check: Reported Beliefs vs Bayesian Posteriors

Table 1: Correlation of Beliefs with Bayesian Posteriors

	Dependent variable:
	Reported Belief
Constant	0.132***
	(0.010)
Bayesian Posterior	0.724***
	(0.019)
Observations	6,720
Adjusted R^2	0.508
Note:	*p<0.1; **p<0.05; ***p<0.01
	SEs clustered by subject.

Back

How do people react to uncertain information signals?

Method:

- Estimate inference and base-rate use (Benjamin, 2019).

How do people react to uncertain information signals?

Method:

- Estimate inference and base-rate use (Benjamin, 2019).

Results:

- Significant over-inference ($c \approx 1.4$).
  - Contrary to result of Benjamin (2019): $c < 1$.
  - Potential reason: introduction of information uncertainty.
- Significant base-rate neglect ($d \approx 0.75$).
- Evidence of confirmation bias if signal confirms prior.
- Individual belief updates are noisy.

Methodology Regression

Analysis: Inference and Base-Rate Use

- Use log-likelihood ratios to analyze inference bias (Benjamin, 2019).
- Estimate inference bias and base-rate neglect jointly:

$$ln(\frac{b_t(R|\mathbf{s_1},...,\mathbf{s_t})}{b_t(B|\mathbf{s_1},...,\mathbf{s_t})}) = \alpha + \beta_1 \cdot ln(\frac{p(\mathbf{s_t}|R)}{p(\mathbf{s_t}|B)}) + \beta_2 \cdot ln(\frac{p_t(R)}{p_t(B)}) + \eta_t$$

  - $b_t(\cdot)$ is a reported belief, and
  - $p(s|R)$ is the probability of seeing s given true state R.

- Interpretation:
  - $\beta_1 = 1$ indicates perfect Bayesian inference.
  - $\beta_1 = 0$ indicates no updating at all.
  - $\beta_2 = 1$ indicates no base-rate neglect.
  - $\beta_2 = 0$ indicates full base-rate neglect.

Back

Robustness Checks - Experimental Design

Anchoring:

- Do not show previously reported belief.
- Finding: No significant influence on updating with retractions or regular updating.
- Other: Too low belief of people that previously under-reacted no longer significant. However, not enough power to find any effect of anchoring.

Backward revision of beliefs:

- Do not show entire history of signals, only previous belief.
- Findings: No significant influence on updating with retractions or regular updating.

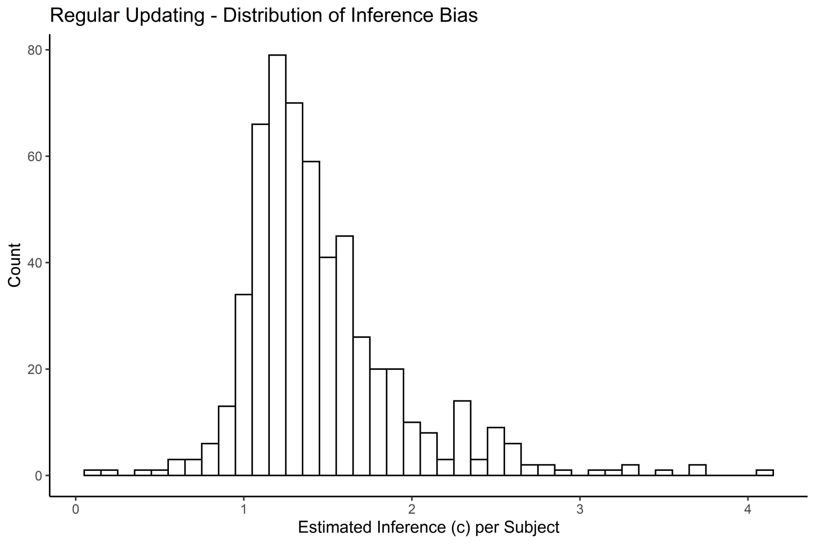
Regular Updating - Inference and Base-Rate Use

Table 2: Updating with Uncertain Signals

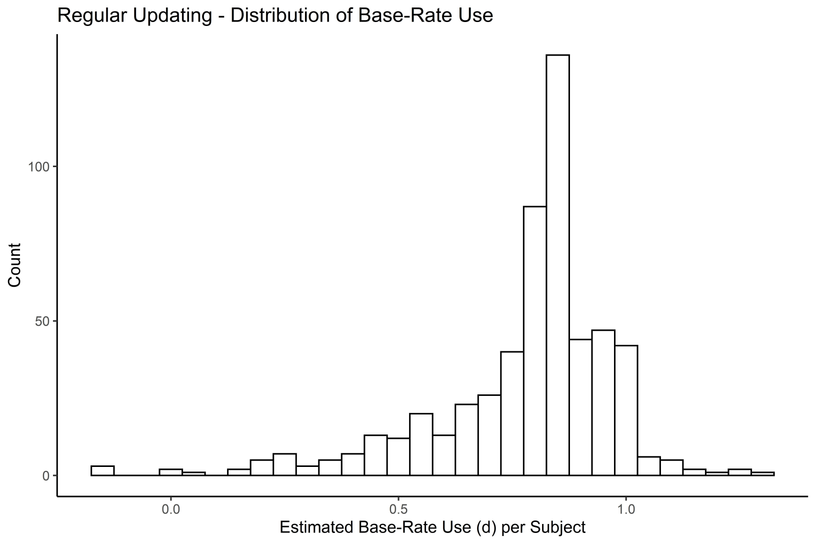
	Dependent variable:		
	Observed Log-Posterior-Ratio		
	OLS	Linear Mixed Effects	
	(1)	(2)	(3)
Constant	-0.034	-0.032	-0.032
	(0.025)	(0.022)	(0.022)
Signal	1.516^{***}	1.505^{***}	1.344^{***}
	(0.060)	(0.060)	(0.079)
Prior	0.704^{***}	0.739^{***}	0.701^{***}
	(0.032)	(0.022)	(0.024)
Signal Confirms Prior			0.425^{***}
			(0.133)
Observations	5,040	5,040	5,040
Adjusted R^2	0.493		
Akaike Inf. Crit.		18,674.110	18,667.390
Note:			*p<0.1; **p<0.05; ***p<0.01
			SEs clustered by subject.

Back

Regular Updating Types - Inference



Regular Updating Types - Base-Rate Use



Analysis: Compressed histories

- Method introduced by (Goncalves et al., 2022).
- A compressed history is given by the exact sequence of signals minus the retracted signal.
- Allows for a clean comparison between people who have seen the same sequence with and without a retraction.
- We estimate: $b_t = \alpha + \beta_1 \cdot r_t + F_{H(R)} + F_{C(H_t)} + \epsilon_t$
  - $H(R)$ refers to the number of seen retractions. Example: RBB would be one red retraction and 2 blue retractions in that order.
  - $C(H_t)$ refers to the compressed history of signals H_t.
  - $F(\cdot)$ denotes the fixed effects for each.
- Interpretation: A positive coefficient β for any combinations of red retracted balls indicates continued influence of retracted signals and vice versa.
- Goncalves et al. (2022) find $\beta > 0$ for a single red retraction.

Back

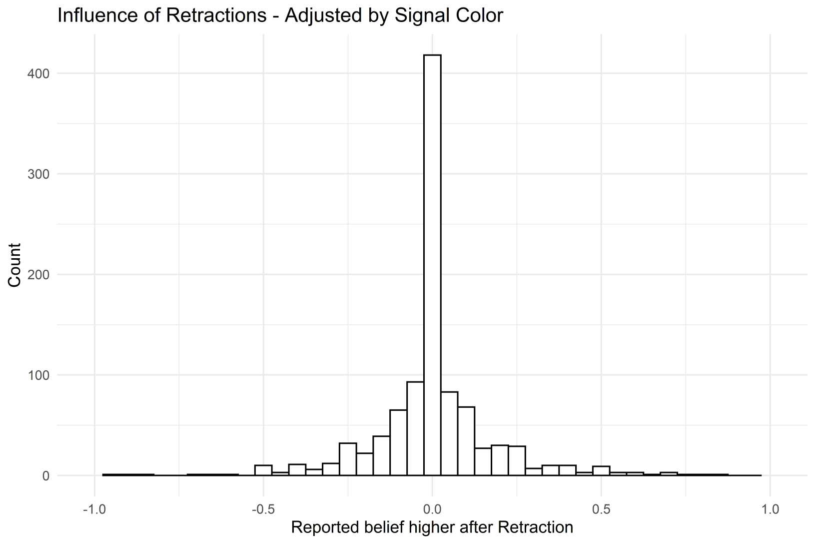
Regression: Initial update explains retraction updating

Table 3: Impact of Retractions on Beliefs

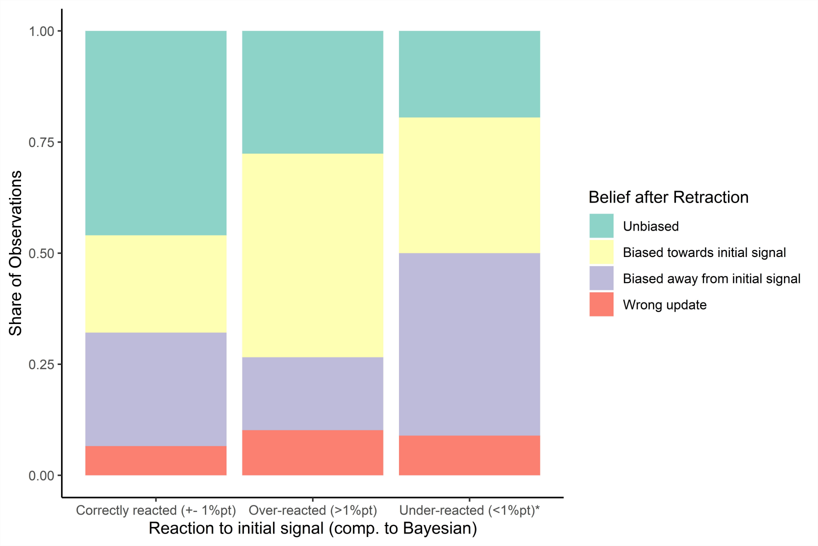
	Dependent variable:
	Belief minus Bay. Post.
Constant	-0.005
	(0.005)
Belief minus Bay. Post. Previously	0.634^{***}
	(0.053)
Observations	1,015
Adjusted R^2	0.325
Note:	*p<0.1; **p<0.05; ***p<0.01
	SEs clustered by subject.

Back

Impact of Retractions - Individual Differences



Impact of Retractions - Categorized by Number of Observations



Impact of Retractions - Compressed History Analysis

Impact of Retractions

	Dependent variable:		
	Reported Belief		
	All histories	All histories	Excluding confirmation histories
	(1)	(2)	(3)
Retraction	-0.011	-0.008	-0.014
	(0.009)	(0.009)	(0.010)
Retraction History: R	0.006	0.003	0.005
	(0.010)	(0.011)	(0.012)
Retraction History: B	-0.003	-0.008	-0.001
	(0.009)	(0.011)	(0.012)
Retraction History: RR	0.004	-0.007	0.003
	(0.014)	(0.016)	(0.020)
Retraction History: BB	-0.017	-0.029^{*}	-0.005
	(0.013)	(0.016)	(0.018)
Retraction History: RB	0.025^{*}	0.014	0.011
	(0.014)	(0.017)	(0.020)
Retraction History: BR	0.001	-0.011	0.022
	(0.014)	(0.017)	(0.021)
Retraction History: RRR	0.053^{**}	0.035	0.054^{***}
	(0.021)	(0.025)	(0.021)
Retraction History: BBB	-0.035	-0.055^{**}	-0.033
	(0.024)	(0.028)	(0.024)
Retraction History: RRB	0.074^{*}	0.057	0.075^{*}
	(0.041)	(0.043)	(0.039)
Retraction History: BBR	0.059^{***}	0.039	0.061^{***}
	(0.023)	(0.027)	(0.022)
Retraction History: RBB	0.030	0.011	0.031
	(0.026)	(0.029)	(0.025)
Retraction History: BRR	0.043	0.023	0.045^{*}
	(0.027)	(0.030)	(0.026)
Retraction History: RBR	0.021	0.001	0.021
	(0.028)	(0.031)	(0.028)
Retraction History: BRB	0.060^{**}	0.041	0.062^{**}
	(0.030)	(0.033)	(0.029)
Compressed History FEs?	Yes	Yes	Yes
Round FEs?	No	Yes	No
Observations	6,660	6,660	3,765
Adjusted R^2	0.498	0.497	0.396
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Are retractions different to 'regular' signals?

Method:

- Retraction of previous signal = new opposite signal.
- Compare two-round updating of regular signals in all histories with mixed signals.
- Example: Compare belief after signals (b, r) to belief after signals $(b, b$ retracted$)$.

Back

Are retractions different to 'regular' signals?

Method:

- Retraction of previous signal = new opposite signal.
- Compare two-round updating of regular signals in all histories with mixed signals.
- Example: Compare belief after signals (b, r) to belief after signals $(b, b$ retracted$)$.

Findings:

- Significant differences between retractions and 'regular' signals.
- On average, people over-react to opposite colored new signal (while no mistake with retractions).

Figure

Back

Alternative Mechanisms

Subject types or initial reaction?

- Subjects can be categorized into types based on their reaction to uncertain signals.
- Subject types do not predict the reaction to a retraction. More

Back

Alternative Mechanisms

Subject types or initial reaction?

- Subjects can be categorized into types based on their reaction to uncertain signals.
- Subject types do not predict the reaction to a retraction.

More

Does anchoring explain the finding?

- Treatment that does not display previously reported belief.
- No significant differences to main treatment.

Regression

Back

Alternative Mechanisms

Subject types or initial reaction?

- Subjects can be categorized into types based on their reaction to uncertain signals.
- Subject types do not predict the reaction to a retraction.

More

Does anchoring explain the finding?

- Treatment that does not display previously reported belief.
- No significant differences to main treatment.

Regression

Are people trying to correct a prior mistake?

- Prior mistake: Misreported belief prior to uncertain signal.
- No, if anything the opposite is true.

Regression

Back

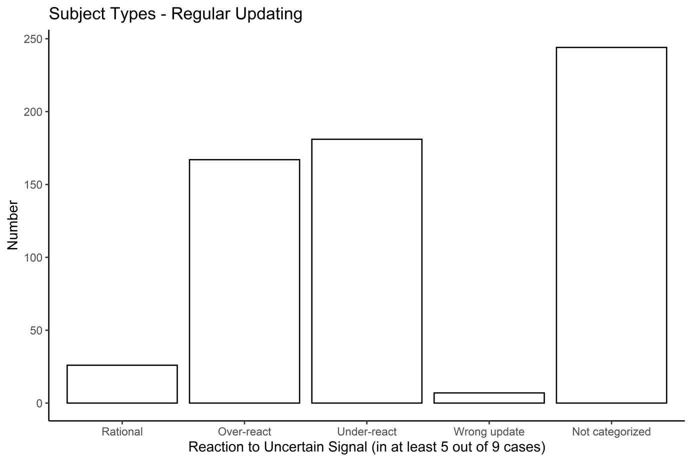
Subject Types

Table 4: Impact of Retractions on Beliefs

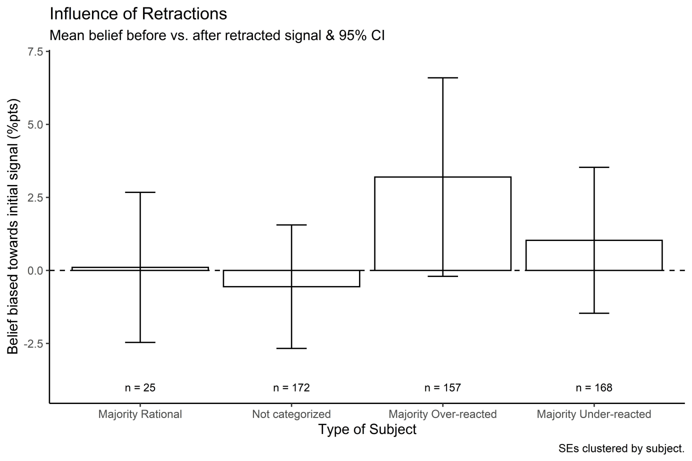
	Dependent variable:
	Belief biased towards initial signal
Constant	0.001
	(0.015)
Type: Not categorized	−0.007
	(0.018)
Type: Majority Over-reported	0.031
	(0.021)
Type: Majority Under-reported	0.009
	(0.018)
Type: Majority Wrong	−0.035
	(0.118)
Observations	1,015
Adjusted R^2	0.002
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
	SEs clustered by subject.

Back

Subject Types



Subject Types



Are reactions to retractions biased to offset previous mistake?

Table 5: Impact of Retractions on Beliefs

	Dependent variable:
	Belief biased towards initial signal
Constant	-0.006
	(0.006)
Initial belief over-report (t-1)	0.699^{***}
	(0.063)
No anchor treatment	0.012
	(0.012)
No anchor treat * initial belief over-report (t-1)	-0.126
	(0.111)
Belief over-report before (t-2)	-0.161^{***}
	(0.044)
Observations	1,015
Adjusted R^2	0.348
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ SEs clustered by subject.

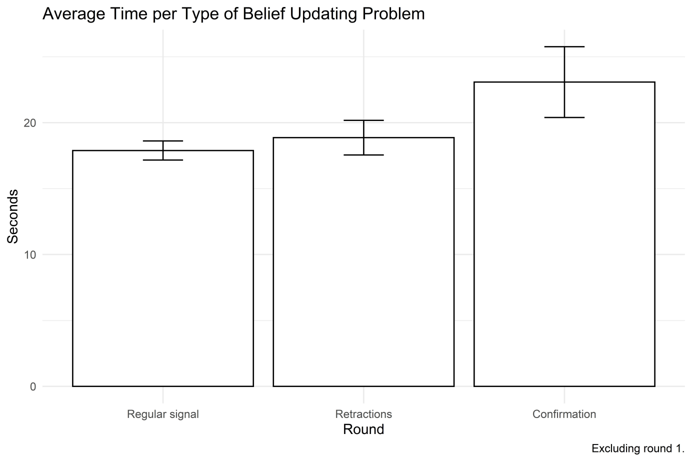
Back

Decision Time by Updating Problem

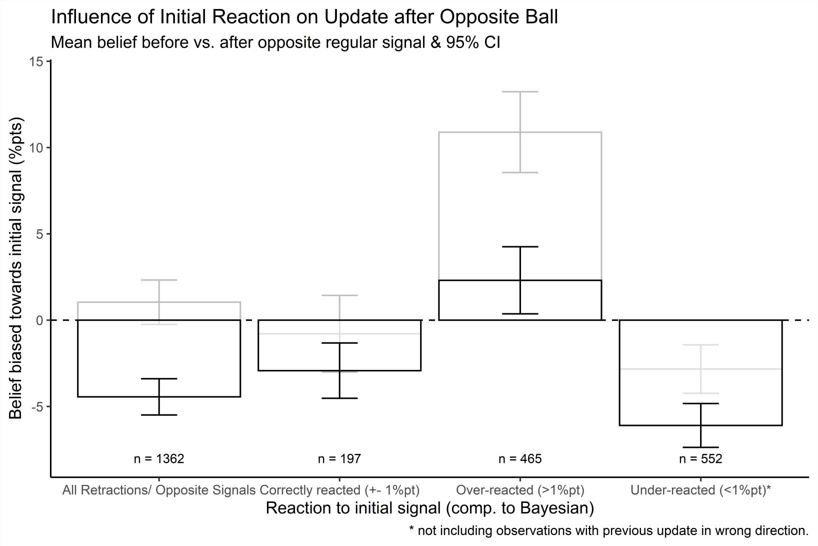
Are retractions more difficult to process?

- Measured by decision time.
- Mechanism suggested by Goncalves et al. (2022).
- No significant difference in decision time in our sample. Graph

Decision Time by Updating Problem



Retractions are different to 'regular' signals



Back

Impact of Verifications on Regular Updating

Updating with Regular Signals

	Dependent variable:		
	Observed Log-Posterior-Ratio		
	(1)	(2)	(3)
Constant	-0.046**	-0.045**	-0.044**
	(0.022)	(0.022)	(0.022)
Signal	1.344***	1.342***	1.330***
	(0.130)	(0.130)	(0.130)
Prior	0.919***	0.920***	0.905***
	(0.066)	(0.066)	(0.069)
Prior * |0.5 - Prior|	-1.047***	-1.048***	-1.010***
	(0.151)	(0.151)	(0.155)
Prior * Round	0.034***	0.034***	0.033***
	(0.003)	(0.003)	(0.003)
Signal * Round	0.056	0.057	0.063*
	(0.038)	(0.038)	(0.038)
Signal * # Previously Verified Signals	-0.182*		
	(0.109)		
Signal * # Previous Retractions		-0.140	
		(0.116)	
Signal * # Previous Confirmations		-0.261**	
		(0.131)	
Signal * # Previous Same Retractions			0.031
			(0.127)
Signal * # Previous Same Confirmations			-0.254
			(0.157)
Signal * # Previous Other Retractions			-0.358***
			(0.132)
Signal * # Previous Other Confirmations			-0.302*
			(0.161)
Observations	4,995	4,995	4,995
Log Likelihood	-9,057.651	-9,058.339	-9,055.329
Akaike Inf. Crit.	18,143.300	18,146.680	18,144.660
Bayesian Inf. Crit.	18,234.530	18,244.420	18,255.430
Note:			*p<0.1; **p<0.05; ***p<0.01

Back

Analysing Confirmations

Rational Posterior

- Simply 'forget' the initial uncertain signal.
- Update knowing signal is informative, using initial prior belief.

Analysis:

- Difference to Bayesian beliefs after confirmation signal.
- Control for influence of initial reaction to uncertain signal.

How do people react to confirmations?

Results:

- Slight under-reaction for Bayesian initial reports.
- As with retractions, initial update explains reaction to confirmation.
  - Also more expected as confirmations change rational beliefs in the same direction as the initial signal.
- Only small differences to regular updating.

Confirmations are similar to ’regular’ signals