Belief Updating and Misinformation

IFREE Presentation

Elias Tsakas, Martin Strobel and **Lars Wittrock**

May 13, 2022

Maastricht University

Introduction

Motivation

- Information is often not 100% reliable.
    - 37% of people reported to come across 'fake news' on a daily basis (European Commission, 2018).
    - This leads to information uncertainty more generally.

Motivation

- Information is often not 100% reliable.
  - 37% of people reported to come across 'fake news' on a daily basis (European Commission, 2018).
  - This leads to information uncertainty more generally.

- Uncertain information frequently gets verified.
  - A verification can lead to confirmation or retraction of previous information.
  - Both of these outcomes are common, for example in the context of scientific research during the Covid-19 pandemic.
  - At least 224 published studies on Covid-19 have been retracted since 2020 ('RetractionWatch.com').

Motivation

- It is unclear how people deal with *information about information*, i.e. verifications.
  - Some evidence that people do not fully 'unlearn' if a previous piece of information is retracted.
    - Continued influence effect - psychology literature
    - Refuted Covid-19 statements, child vaccinations and autism, deceptive mouthwash advertising[1], ...
  - It is unclear how people react to confirmations and how verifications affect future updating.

- A clear understanding of how people respond to information has important implications for the communication of policies, especially if these are based on uncertain information.

[1] Examples mentioned in Meyer et al. (2020) and Lewandowsky et al. (2012)

Literature

- Belief updating problems first studied in 60s and 70s.
  - Phillips and Edwards (1966); Kahneman and Tversky (1971,1974); and many others.
- Benjamin (2019) – Meta study
  - Strong evidence of under-inference in general.
  - No pooling of data with sequential updating.
  - Recency and primacy effect with sequential updating.
- Goncalves et al. (2021)
  - Subjects under-infer from retractions.
  - Beliefs are more sensitive to new signals after a retraction.
- Psychology literature - Continued Influence Effect[2]
  - Mechanism is based on memory and recall.
  - Works best (or mainly?) with narratives that have a causal structure.
  - Cognitive ability partly explains the size of the effect.

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

Research Questions

1. How do people update their belief in response to being told a previous signal was (un)informative?
   - Retraction: CIE in psychology, confirmed by Goncalves et al. (2021).
   - Confirmation: not tested. [Not part of today]

2. When do people under-react or over-react to a retraction?

3. Do verifications in the past affect inference from regular signals later on?

Experimental Design

Experimental Design

Ball and urn framework:

- 2 urns filled with different distributions of balls. One urn is randomly picked.
- Task: guess which of the urns was selected after receiving hints from the selected urn.
- Deliberately abstract design to avoid context dependencies. More

Experimental Design

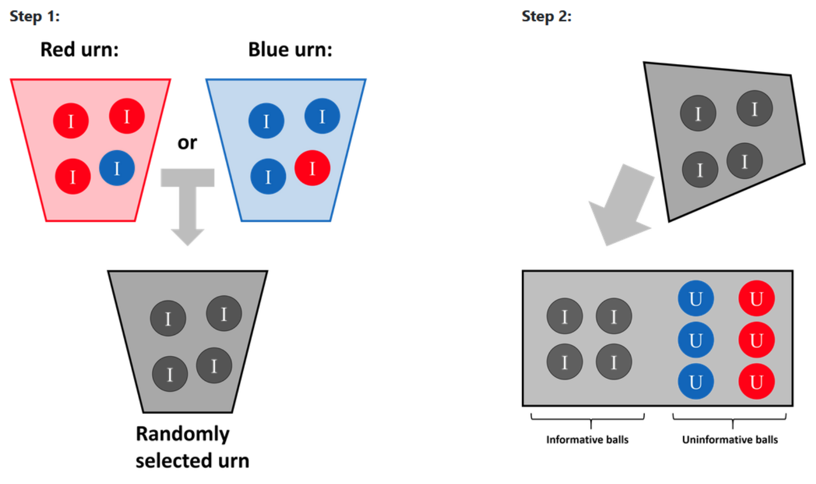
Ball and urn framework:

- 2 urns filled with different distributions of balls. One urn is randomly picked.
- Task: guess which of the urns was selected after receiving hints from the selected urn.
- Deliberately abstract design to avoid context dependencies. More

Modification:

- 4 balls from the selected urn are put into a box with 6 other, irrelevant balls, i.e. 60% chance of a signal being uninformative.
- 12 rounds of information. 9 balls are shown and 3 times information about the previous ball is displayed.
- The order of verification rounds varies per subject.

Urns and Black Box



Participant 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 +

%

Implementation

- Experiment was conducted online with subjects recruited from Prolific.
- Hypotheses and analysis were pre-registered on 'aspredicted.org'.
- You can have a look at the complete experiment by visiting: `https://updatingsurvey.herokuapp.com/join/kiripame`.

Analysis

Analysis

Defining posteriors:

- Regular: Posterior is given by Bayes' rule, using the previous belief as the prior.
- Retractions: Belief before the retracted uninformative signal.
- Confirmations: Multiple options exist.

Measuring bias:

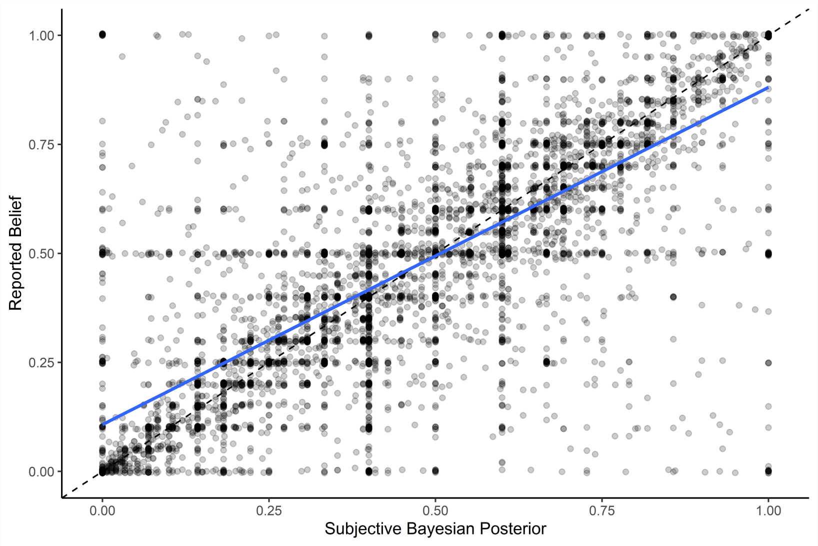
- Regular: Estimate inference and base-rate use. Method
- Retractions:
  - Estimate inference and base-rate use.
  - Average belief difference before and after retracted signal.
  - Compare compressed histories. Method
- Confirmations: Estimate inference and base-rate use.

Results

Sample Overview

- 606 subjects completed the experiment.
- 48 were removed as outliers (pre-registered criteria).
- In total 6,696 observations.
- Median time to complete survey 17 minutes.
- Average payoff is 4.80€.
- Equal ratio male/female.

Reported Beliefs vs Posteriors



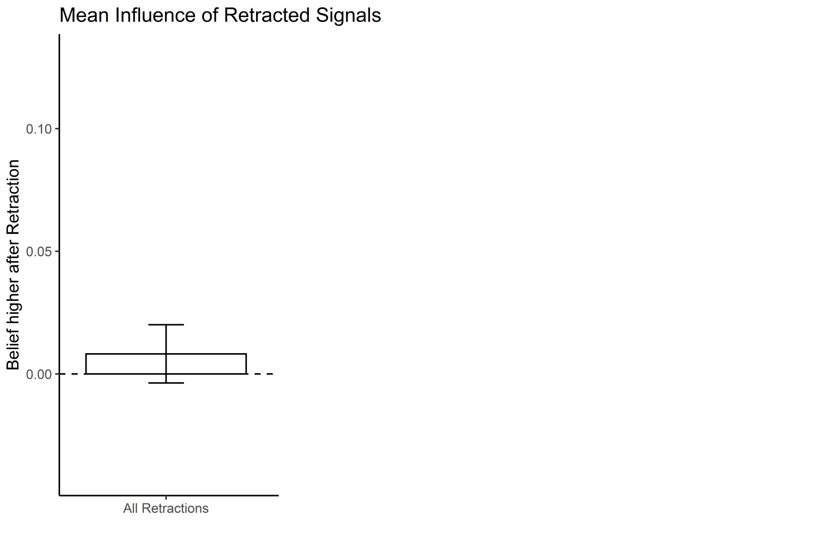
Regression

How do people react to retractions?

In general:

- No significant differences in beliefs before and after seeing an uninformative signal.
- Confirmed by analyzing compressed histories. Regression
- Contrary to Goncalves et al. (2021).

Retractions do not bias subjects on average



How do people react to retractions?

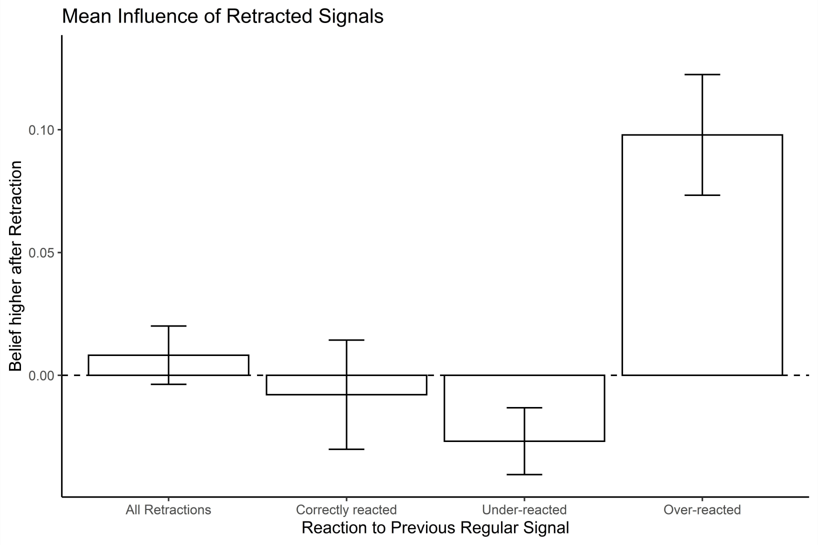
In general:

- No significant differences in beliefs before and after seeing an uninformative signal.
- Confirmed by analyzing compressed histories. Regression
- Contrary to Goncalves et al. (2021).

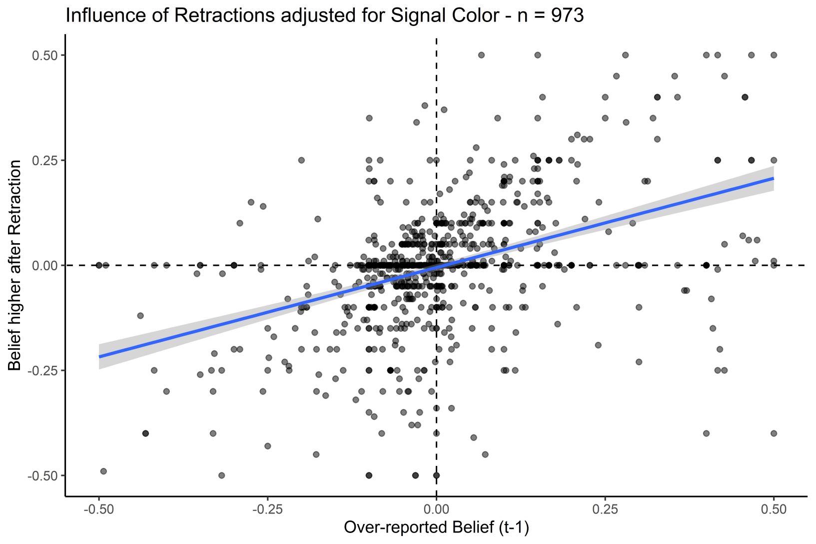
Previous update explains reaction to retractions:

- Previously over-reacted: Too high belief.
- Previously under-reacted: Too low belief.
- Previously correctly reacted: Correct belief.

Previous update explains reaction to retractions



Previous update explains reaction to retractions



Do verifications impact future regular updating?

Regular updating:

Regression

- Significant over-inference ($c \approx 1.4$).
  - Potential reason: introduction of information uncertainty.
- Significant base-rate neglect ($d \approx 0.75$).
- Evidence of confirmation bias.

Do verifications impact future regular updating?

Regular updating: Regression

- Significant over-inference ($c \approx 1.4$).
  - Potential reason: introduction of information uncertainty.
- Significant base-rate neglect ($d \approx 0.75$).
- Evidence of confirmation bias.

Impact of verifications: Regression

- No intuitively plausible finding.
- It seems that the number of past verifications of the other colored ball (retractions or confirmations) significantly decrease inference.
  - Potential reason: gambler's fallacy?

Summary

Summary

Findings:

- Updating with retractions depends on previous reaction (even in a neutral setting!)
- No continued influence on average (contrary to some earlier literature)
- Past verifications have no clear influence on future updating.

Next steps:

- Analysis of updating after confirmations.
- Investigate mechanism further?
- Influence of motivated beliefs?

Questions?

Thank you for your attention!

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.

European Commission (2018). Fake news and disinformation online. Technical report.

Goncalves, D., Libgober, J., and Willis, J. (2021). Learning versus unlearning: An experiment on retractions.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3):106–131.

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

Appendix

Why an Abstract Experimental Setting?

- An abstract setting avoids context dependencies. It seems plausible that settings with motivated beliefs would strengthen the effect.
- Our setting allows a calculation of 'correct' beliefs, using Bayes' rule. The Bayesian framework is a commonly used benchmark to estimate the inference bias.
- Retractions and confirmations of previous signals are unambiguous. In reality this is often more difficult.
- Findings can be easily compared to the broader literature on belief updating.
- Belief elicitation can be incentivized.

Back

Analysis: Inference and Base-Rate Use

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

$$ln\left(\frac{r_t(R|s_1,...,s_t)}{1-r_t(R|s_1,...,s_t)}\right) = \beta_0 + \beta_1 \cdot ln\left(\frac{p(s_t|R)}{p(s_t|B)}\right) + \beta_2 \cdot ln\left(\frac{r_{t-1}(R|s_1,...,s_{t-1})}{1-r_{t-1}(R|s_1,...,s_{t-1})}\right) + \eta_t$$

  - $r_t(R|s_t)$ is the reported belief of R given signals s_1 until s_t in round t.
  - $p(s|R)$ 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

Analysis: Compressed histories

- Method introduced by (Goncalves et al., 2021).
- 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: $r_t(R|s_1, ..., s_t) = \beta_0 + \beta \cdot F_{ret_t} + F_{C(H_t)} + \eta$
  - ret_t 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. (2021) find $\beta > 0$ for a single red retraction.

Back

Sanity Check

OLS Regression Output

	Dependent variable:
	Reported Belief
Constant	0.111^{***}
	(0.004)
Subjective Posterior	0.765^{***}
	(0.008)
Observations	6,660
R^2	0.569
Adjusted R^2	0.569
Residual Std. Error	0.173 (df = 6658)
F Statistic	$8{,}779.675^{***}$ (df = 1; 6658)
Note:	*p<0.1; **p<0.05; ***p<0.01

Back

Robustness Checks - Experimental Design

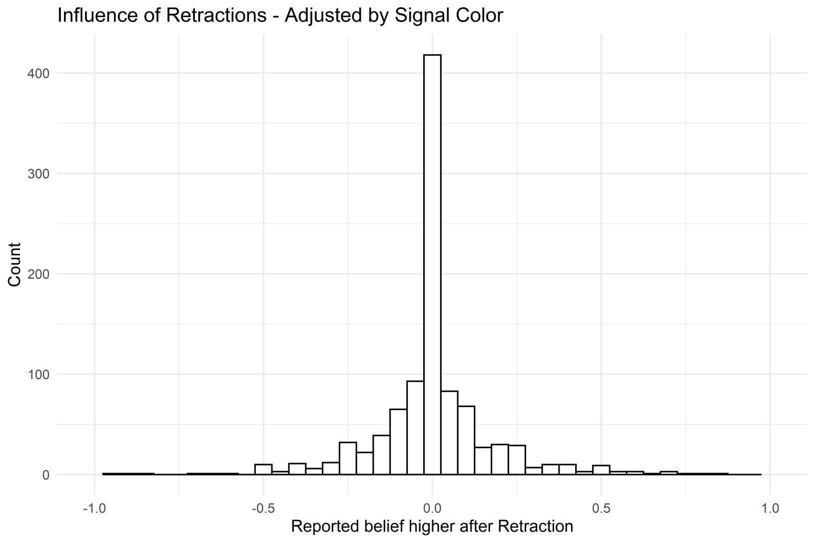
Anchoring:

- 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 no show entire history of signals, only previous belief.
- Findings: No significant influence on updating with retractions or regular updating.

Impact of Retractions - Individual Differences



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:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Regular Updating - Inference and Base-Rate Use

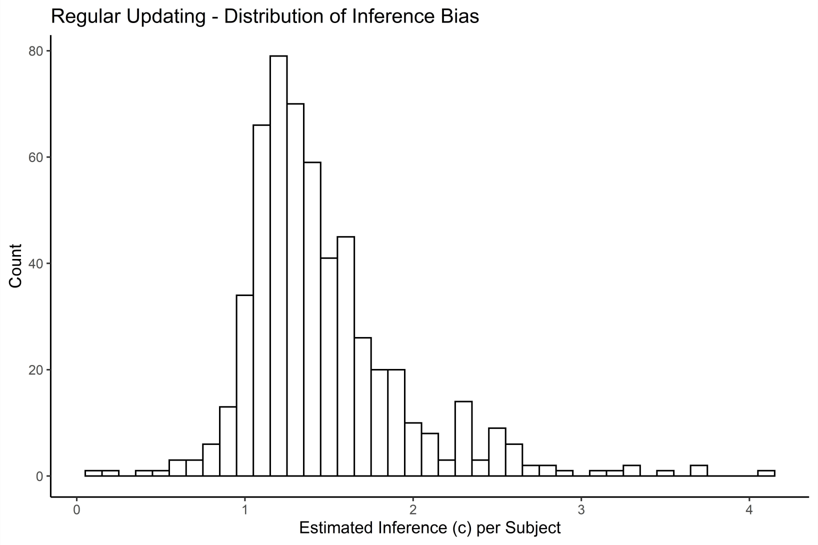
Updating with Regular Signals

	Dependent variable:			
	Observed Log-Posterior-Ratio			
	OLS		*linear mixed-effects*	
	(1)	(2)	(3)	(4)
Constant	-0.047^{**}	-0.046^{**}	-0.039^{*}	-0.042^{*}
	(0.023)	(0.023)	(0.022)	(0.022)
Signal	1.447^{***}	1.260^{***}	1.461^{***}	1.313^{***}
	(0.059)	(0.077)	(0.062)	(0.080)
Prior	0.743^{***}	0.729^{***}	0.781^{***}	0.749^{***}
	(0.010)	(0.011)	(0.020)	(0.022)
Signal Confirms Prior		0.490^{***}		0.384^{***}
		(0.128)		(0.129)
Observations	4,995	4,995	4,995	4,995
R^2	0.539	0.540		
Adjusted R^2	0.539	0.540		
Log Likelihood			-9,103.832	-9,101.223
Akaike Inf. Crit.			18,227.660	18,224.450
Bayesian Inf. Crit.			18,292.830	18,296.120
Residual Std. Error	1.618 (df = 4992)	1.616 (df = 4991)		
F Statistic	$2{,}919.235^{***}$ (df = 2; 4992)	$1{,}956.286^{***}$ (df = 3; 4991)		

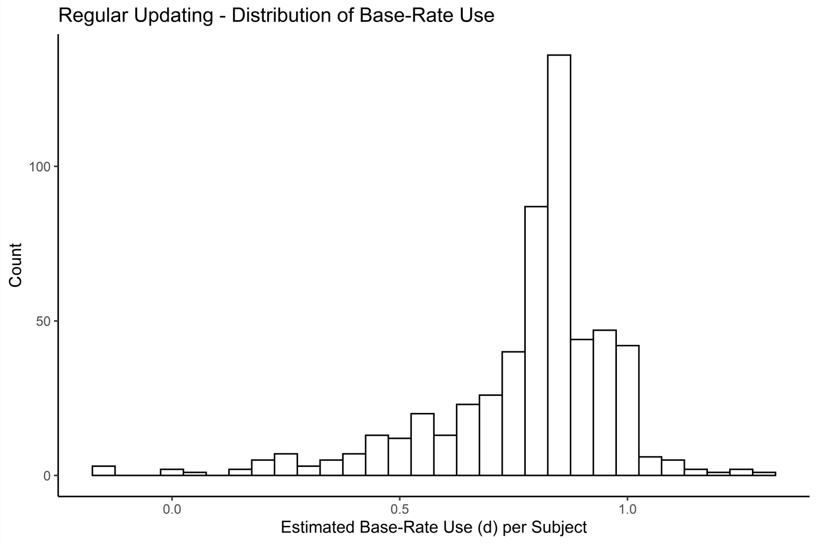
Note: *p<0.1; **p<0.05; ***p<0.01

Back

Regular Updating Types - Inference



Regular Updating Types - Base-Rate Use



Impact of Verifications on Regular Updatring

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