Title: Supplementary analyses appendix for "Gamified inoculation reduces susceptibility to misinformation from political ingroups"
Authors: Cecilie Steenbuch Traberg (1), Jon Roozenbeek (1,2), Sander van der Linden (1)
Date: April 30th, 2024
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix: Supplementary analyses

To test whether playing the *Bad News* game reduces the perceived reliability of reliable/true information ("real news"), we conducted a Bayesian paired-samples *t*-test on the pre- and post-inoculation reliability judgments of true news headlines. Doing so shows that the perceived reliability of true news is reduced post-inoculation compared to pre-inoculation (BF_{10} = 4.359, indicating moderate support for the alternative hypothesis). However, we note that this effect is very small (δ = .119, which represents the population-level version of Cohen's *d*). Furthermore, an exploratory Bayesian paired-samples *t*-test on the pre- and post-scores for the difference between average reliability ratings for misinformation and true news (i.e., veracity discernment) yields strong support for the alternative hypothesis (BF_{10} = 2.49*1013, δ = .332), meaning that inoculation strongly reduces the perceived reliability of misinformation relative to true news. See Table S2 and Figure S1 for a full overview.

Using a frequentist approach, a paired-samples *t*-test showed that the difference between perceived reliability of factual headlines pre- and post-intervention, $t(1303) = 4.31$, $M_{diff} = 0.13$, $SE = 0.03$, $p < .001$, but only at a small effect size of $d = 0.12$.

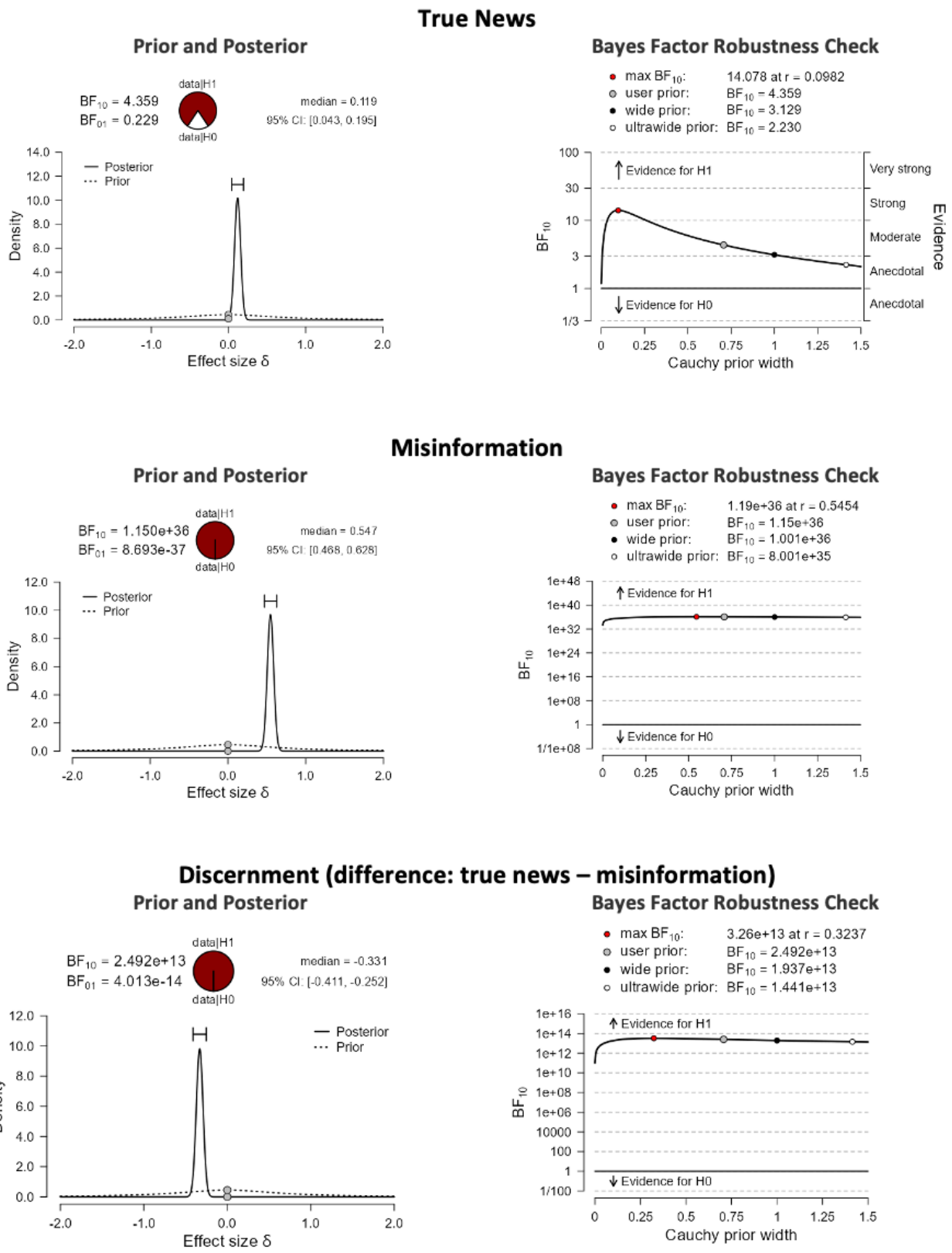

Figure S1. Bayes factors and effect sizes (left panels) and robustness checks with various priors and levels of support for the alternative hypothesis H1 that there is a difference in perceived reliability when comparing pre- and post-gameplay scores (right panels) for true news, misinformation, and veracity discernment.

Table S1A. *Bayesian ANOVA results for the interaction between experimental condition (group) and time (pre-post intervention): Model comparison – PerceivedReliability (perceived reliability of misinformation).*

Models	P(M)	P(M|data)	BF_M	BF_{10}	Error %
Null model (incl. Session.ID)	0.2	3.59E-39	1.44E-38	1	
Group	0.2	8.21E-37	3.28E-36	228.86	1
Time	0.2	0.00362	0.0145	1.01E+36	1.16
Group + time	0.2	0.95949	94.7437	2.67E+38	2.85
Group + time + Group ✻ time	0.2	0.03689	0.1532	1.03E+37	1.57

*Note: All models include Session.ID. P(M) represents the prior probabilities of each model; since we have five models, this is 1/5 = 0.2 for each model. P(M|data) signifies the posterior probability of each model after seeing the data. BF_m is a Bayes Factor that compares each model to the P(M|data) of the other models. BF_{10} gives the probability of the data if H1 is true, P(D|H1). The inclusion Bayes Factor ($BF_{Inclusion}$) represents the evidence in the data for including a predictor (the interaction group * time).*

Table S1B. *Bayesian ANOVA results for the interaction between experimental condition (group) and time (pre-post intervention): Analysis of effects – PerceivedReliability.*

Analysis of Effects – PerceivedReliability		
P(incl)	P(incl|data)	$BF_{Inclusion}$
0.6	0.9964	183.583
0.6	1	6.00E+15
0.2	0.0369	**0.153**

Note: Analysis was conducted using the "JSQ" package in Jamovi (https://www.jamovi.org). Specifically, we ran a Bayesian ANOVA with the perceived reliability of misinformation (PerceivedReliability) as the dependent variable, experimental condition (group) and time (pre-post) as fixed factors, and session.ID (i.e., participant ID) as a random effect. We used the long-format dataset available on the Harvard Dataverse.

Table S2. *Paired-samples Bayesian t-tests for the pre-post difference in the perceived reliability of true news (FactsPre/FactsPost), misinformation (FakePre/FakePost), and the difference between ratings of true news and misinformation, i.e., "veracity discernment" (DiffPre/DiffPost). Bottom table shows descriptive statistics.*

			BF_{10}	Error %
FactsPre	-	FactsPost	4.36	0.00547
FakePre	-	FakePost	1.15E+36	4.71E-43
DiffPre	-	DiffPost	2.49E+13	7.82E-20

Descriptives

	N	*Mean*	*SD*	*SE*
FactsPre	657	4.98	1.14	0.0446
FactsPost	652	4.84	1.37	0.0538
FakePre	657	3.16	1.16	0.0451
FakePost	657	2.55	1.38	0.054
DiffPre	657	1.82	1.54	0.0601
DiffPost	652	2.29	1.76	0.0689

Table S3. *Sample composition in N (count) and percentages of total sample.*

Category	Groups
Education	High school or less (107; 16.3%) Higher degree (288; 43.8%) Some college (262; 39.9%)
Gender	Male (337; 51.3%) Female (284; 43.2%) Other (36; 5.5%)
Age	18-29 (464; 70.6%) 30-49 (156; 23.7%) Over 50 (37; 5.6%)
Political ideology	1 (67; 10.2%) 2 (211; 32.1%) 3 (242; 36.8%) 5 (75; 11.4%) 6 (25; 3.8%) 7 (37; 5.6%)

Note: Political ideology is coded from very liberal (1) to very conservative (7).