Supplementary Materials: Pilot Study Results for

Do Registered Reports Make Scientific Findings More Believable to the Public?

Elaine Costa

University of Utah, Salt Lake City, UT, USA

Yoel Inbar

University of Toronto, ON, Canada

David Tannenbaum

University of Utah, Salt Lake City, UT, USA

Author Note:

Elaine Costa, Department of Management, University of Utah. Yoel Inbar, Department of Psychology, University of Toronto. David Tannenbaum, Department of Management, University of Utah. Correspondence concerning this article should be addressed to Elaine Costa, email: elaine.costa@utah.edu

Supplementary Materials: Pilot Study Results

Table S1

Means and Standard Deviations by Condition

	Registered Report				Non-Registered Report			
	Plausible		Implausible		Plausible		Implausible	
Scientific Finding	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Atheists and agnostics were [less/no more biased] against gender fluid individuals than those with a religious belief system.	5.19	1.45	4.86	1.33	5.48	1.05	4.51	1.35
Couples with children are [just as likely/more likely] to cheat.	5.04	1.21	4.32	1.56	4.99	1.35	4.26	1.35
Compared to non-prisoners, ex-prisoners are [no less/less] reliable and trustworthy at work.	5.04	1.36	4.49	1.39	4.84	1.34	4.40	1.35
Fake news [helped/did not help] elect Donald Trump.	4.82	1.50	4.15	1.61	4.45	1.58	3.97	1.62
Gender discrimination was [found/not found] in small business lending.	5.25	1.26	4.22	1.52	5.27	1.15	4.28	1.48
Marijuana users were [not any more likely/more likely] than non-users to use other illegal drugs like cocaine or heroin.	5.45	1.21	4.66	1.54	5.24	1.36	4.57	1.59
Raising the minimum wage [does not lead/leads] to increased unemployment.	5.22	1.11	4.34	1.51	5.15	1.42	4.14	1.68
Greater social media use [leads/does not lead] to depression and reduced well-being.	5.45	1.17	4.20	1.54	5.38	1.38	3.88	1.48
Math and science scores were found on average to be [the same/lower] for female students.	5.29	1.45	4.15	1.45	5.27	1.17	4.08	1.50
Playing violent video games [has no effect on/increases] aggressive behavior in adolescents.	5.24	1.45	4.32	1.67	5.12	1.46	4.36	1.62

Note: Description of scientific findings are abridged versions of the ones viewed by participants, which also included a brief topic introduction and methodological description of the study. Credibility scores of study findings were rated on 7-point scales (1 = strongly disagree, 7 = strongly agree). The first outcome listed in brackets for each finding represents the plausible condition, while the second represents the implausible condition.

Pilot Study Results (OLS Models)

Table S2

Effect of Registered Reports and Scientific Bias on Scientific Findings' Credibility

	(1)	(2)	(3)	(4)	(5)	(6)
Study Outcome	0.849***	0.872***		0.818***		0.867***
	(0.055)	(0.071)		(0.071)		(0.092)
Registered Report	0.098*	0.120			0.130*	0.164
	(0.048)	(0.071)			(0.066)	(0.094)
Scientific Bias			-0.179**	-0.180*	-0.148*	-0.142
			(0.059)	(0.084)	(0.075)	(0.111)
Registered Report x Study Outcome		-0.044				-0.104
		(0.088)				(0.115)
Scientific Bias x Study Outcome				0.065		0.008
				(0.112)		(0.146)
Scientific Bias x Registered Reports					-0.064	-0.086
					(0.101)	(0.142)
Scientific Bias x Study Outcome x Registered Reports						0.123
						(0.179)
Intercept	4.543***	4.531***	5.094***	4.671***	5.028***	4.590***
	(0.074)	(0.081)	(0.072)	(0.078)	(0.079)	(0.094)
Vignette Fixed Effects	yes	yes	yes	yes	yes	yes
Observations	4,000	4,000	4,000	4,000	4,000	4,000
Participants	800	800	800	800	800	800
Vignettes	10	10	10	10	10	10
R^2	0.095	0.095	0.017	0.096	0.019	0.098

Note: Columns correspond to OLS regression coefficients, with participant-clustered standard errors in parentheses. The dependent variable in both models is an index of credibility judgments scored on a 7-point scale, with positive values denoting higher credibility judgments. Study Outcome takes on the value of 0 if the scientific finding was rated as implausible and 1 if the scientific finding was rated as plausible, based on pilot data. Registered Report takes on the value of 1 for the presence of a registered report, and 0 for a non-registered report. Scientific Bias takes on the value of 1 for the presence of scientific bias beliefs and 0 for its absence. For scenarios, we dummy-coded 10 vignettes with the "Atheists/Agnostics" scenario representing the reference value. Significance levels: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

Pilot Study Results (Models Using Cross-Random Effects)

Table S3

Effect of Registered Reports and Scientific Bias on Scientific Findings' Credibility

	(1)	(2)	(3)	(4)	(5)	(6)
Study Outcome	0.864***	0.893***		0.832***		0.881***
	(0.096)	(0.106)		(0.102)		(0.117)
Registered Report	0.099*	0.128*			0.120*	0.169*
	(0.044)	(0.062)			(0.057)	(0.082)
Scientific Bias			-0.149*	-0.181*	-0.130	-0.143
			(0.058)	(0.073)	(0.074)	(0.097)
Registered Report x Study Outcome		-0.056				-0.101
		(0.087)				(0.114)
Scientific Bias x Study Outcome				0.069		0.028
				(0.089)		(0.126)
Scientific Bias x Registered Reports					-0.042	-0.086
					(0.089)	(0.125)
Scientific Bias x Study Outcome x Registered Reports						0.093
						(0.177)
Intercept	4.254***	4.240***	4.796***	4.380***	4.737***	4.298***
	(0.074)	(0.077)	(0.111)	(0.077)	(0.115)	(0.087)
Vignette Fixed Effects	yes	yes	yes	yes	yes	yes
Observations	4,000	4,000	4,000	4,000	4,000	4,000
Participants	800	800	800	800	800	800
Vignettes	20	20	20	20	20	20
R^2	0.084	0.084	0.002	0.085	0.004	0.087

Note: Columns correspond to regression coefficients from models with cross-random effects for participants and scenarios. The dependent variable in all models is an index of credibility judgments scored on a 7-point scale, with positive values denoting higher credibility judgments. Study Outcome takes on the value of 0 if the scientific finding was rated as implausible and 1 if plausible based on pilot data. Registered Report takes on the value of 1 for the presence of a registered report, and 0 for a non-registered report. Scientific Bias takes on the value of 1 for the presence of scientific bias beliefs and 0 for its absence. Significance levels: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

Table S4

Effect of Registered Reports and Scientific Bias on Scientific Findings' Credibility (Aggregated Results)

	(1)	(2)	(3)	(4)	(5)	(6)
Study Outcome	0.735***	0.757***		0.689***		0.703***
	(0.031)	(0.040)		(0.040)		(0.052)
Registered Report	0.018	0.040			-0.082	-0.065
	(0.032)	(0.044)			(0.043)	(0.058)
Scientific Bias			-0.197***	-0.231***	-0.316***	-0.356***
			(0.034)	(0.047)	(0.046)	(0.064)
Registered Report x Study Outcome		-0.043				-0.029
		(0.051)				(0.068)
Scientific Bias x Study Outcome				0.101		0.114
				(0.062)		(0.080)
Scientific Bias x Registered Reports					0.236***	0.245**
					(0.066)	(0.087)
Scientific Bias x Study Outcome x Registered Reports						-0.023
						(0.102)
Intercept	4.338***	4.326***	4.796***	4.443***	4.836***	4.474***
	(0.052)	(0.055)	(0.051)	(0.054)	(0.055)	(0.061)
Vignette Fixed Effects	yes	yes	yes	yes	yes	yes
Study Fixed Effects	yes	yes	yes	yes	yes	yes
Observations	11,500	11,500	11,500	11,500	11,500	11,500
Participants	2,300	2,300	2,300	2,300	2,300	2,300
Vignettes	10	10	10	10	10	10
R^2	0.07	0.07	0.01	0.07	0.01	0.08

Note: Columns correspond to OLS regression coefficients aggregating data across the pilot and main study, with participant-clustered standard errors in parentheses. The dependent variable in both models is an index of credibility judgments scored on a 7-point scale, with positive values denoting higher credibility judgments. Study Outcome takes on the value of 0 if the scientific finding was rated as implausible and 1 if the scientific finding was rated as plausible, based on pilot data. Registered Report takes on the value of 1 for the presence of a registered report, and 0 for a non-registered report. Scientific Bias takes on the value of 1 for the presence of scientific bias beliefs and 0 for its absence. For scenarios, we dummy-coded 10 vignettes with the "Atheists/Agnostics" scenario representing the reference value. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Aggregate Analyses Results

In line with our pre-registration, we report the results when aggregating the results from our pilot study and main study. We combined the two datasets (N = 2,300) and re-ran the same OLS regressions as our first and second models from the main paper, while including a fixed effect indicator for study origin (0 = pilot study, 1 = main study). Similar to before, our coefficient of interest for model 1 was the registered report variable, while the coefficient of interest for model 2 was the interaction term.

First looking at model 1 on Table S4, the effect of RRs on credibility ratings was positive but not statistically significant, b = 0.018, SE = 0.032, p = 0.564. Looking at model 2, the interaction term between RRs and study outcome was negative but not significant, b = –0.043, SE = 0.051, p = 0.401. To examine whether the effect of RRs was reliably different across studies, we conducted an analysis with a specification similar to our first model, but now also including an interaction term between study origin and registered reports. The interaction effect for this model was marginally significant, b = –0.121, SE = 0.063, p = 0.056. As suggested by the results in the different studies, average marginal effects indicate a positive and statistically significant effect of registered reports in the pilot, but a negative and not significant effect in the main study.

Moving to our exploratory analysis, we regressed credibility scores onto scientific bias beliefs, registered reports, and an interaction term between the two variables, while including study origin fixed effects. The interaction between RRs and scientific bias is again positive and significant, b = 0.236, SE = 0.066, p < 0.001. An analysis of the simple slopes indicated that while registered reports increased credibility ratings for those who view scientists as biased, b = 0.154, SE = 0.050, p = 0.002, it had a negative but weak effect on credibility ratings among those

who do not view scientists as biased, $b = –0.082$, SE = 0.043, $p = 0.056$. In other words, we find the result of RRs reducing credibility of findings among those who are typically not skeptical of scientists to be inconsistent, and in aggregate, results suggest that most of the gains in credibility judgments of scientific findings conducted under RRs are realized by those who believe scientists are biased.