

**Supplements for “The insights from the crowd: Drawing inferences
from many approaches to key empirical questions in international business”**

Table of Contents

Supplement 1: Deviations from preregistered analysis plan	2
Supplement 2: Control variables used by each analyst	3
Supplement 3: Frequency of IV and DV operationalizations.....	7
Supplement 4: Results selecting experts and specific IVs and DVs	11
Supplement 5: Distribution of beliefs regarding each research question	19
Supplement 6: Beliefs and variance in effect size estimates	23
Supplement 7: Expertise, beliefs, and evidence	25
Supplement 8: Analysts’ academic profiles and demographics.....	26
Supplement 9: Approaches to preventing and detecting strategic and inherent elasticity in analytic procedures	28
References	38

Authors: Andrew Delios, Tianyou Hu, Shu Yu, Nan Zhou, the many analysts, and Eric Uhlmann.
(See the full author list in the published JIBS paper.)

Correspondence to:

Tianyou Hu, Faculty of Business Administration, University of Macau, E22, Avenida da Universidade, Taipa, Macau. Email: tianyouhu@um.edu.mo

Nan Zhou, School of Economics and Management & Advanced Institute of Business, Tongji University, 1500 Siping Road, Shanghai, China, 200092. Email: zhounan38@hotmail.com

Supplement 1: Deviations from preregistered analysis plan

For the final project report, we ran and carried out the following analyses which we had not pre-registered beforehand.

1. Testing the importance of variable operationalizations (IV, DV, and IV-DV pairing) in explaining effect size dispersion. This analysis follows on Schweinsberg et al. (2021), and conceptually replicates the results in that paper.
2. Comparing whether analyst or research question explains more variance in estimates. This analysis follows on Landy et al. (2020), and our findings are directly contrary to the results in that paper.
3. Selecting only experts and examining the range of estimates, direction counts, and testing heterogeneity for RQ1-4. This follows on Silberzahn et al. (2018) and Menkveld et al. (2024) and conceptually replicates both papers.
4. Testing whether expertise moderates the relationship between pre-beliefs and/or post-beliefs and empirical estimates (Supplement 7). This idea occurred to us during the analysis phase; the results were unsupportive of our hypothesis that experts would exhibit less confirmation bias and more rational belief updating.
5. The exploratory analysis selecting only experts and the single most common dependent variable and re-examining the dispersion in estimates (Supplement 4). These results ran contrary to our expectation that we would be able to arrive at tentative directional conclusions for RQ3 and RQ4. However, we were able to find comparative consistency in estimates within some of the IV-DV pairings for RQ2 (see Supplement 4).
6. The tests of heterogeneity in beliefs across the four research questions (Supplement 5)
7. The tests of the relationships between pre-beliefs and post-beliefs for each research question reported in the main manuscript.

Supplement 2: Control variables used by each analyst

Table S2-1. Control variables used by each analyst for Research Question 1

[illegible]

Notes: Analyst numbers are listed on the top and control variables on the left. A shaded box indicates that the corresponding analyst used the control variable in their final model. The table is ordered by the frequency with which each control variable was used.

Table S2-2. Control variables used by each analyst for Research Question 2

[illegible]

Notes: Analyst numbers are listed on the top and control variables on the left. A shaded box indicates that the corresponding analyst used the control variable in their final model. The table is ordered by the frequency with which each control variable was used.

Table S2-3. Control variables used by each analyst for Research Question 3

[illegible]

Notes: Analyst numbers are listed on the top and control variables on the left. A shaded box indicates that the corresponding analyst used the control variable in their final model. The table is ordered by the frequency with which each control variable was used.

Table S2-4. Control variables used by each analyst for Research Question 4

[illegible]

Notes: Analyst numbers are listed on the top and control variables on the left. A shaded box indicates that the corresponding analyst used the control variable in their final model. The table is ordered by the frequency with which each control variable was used.

Supplement 3: Frequency of IV and DV operationalizations

Table S3-1: IV choice and frequency for research question 1

No.	IV Choices	Frequency
1	1=WOS; 2=JV; 3=Acquisition	36
2	Whether JV (1) or WOS/acquisition (0)	20
3	Whether JV (1) or WOS (0)	12
4	Whether JV (1) or acquisition (0)	2
5	Whether WOS (1) or JV/Acquisition (0)	14
6	Whether WOS (1) or acquisition (0)	7
7	Whether acquisition (1) or JV/WOS (0)	18
8	Ownership percentage by largest Japanese firm	5
9	Whether Japanese ownership is larger than 80% or 95%	5
10	Whether it is acquisition of more than 50% of the ownership	2
11	Whether a JV is majority (1) or minority (0) owned	5

Table S3-2: IV choice and frequency for research question 2

No.	IV Choices	Frequency
1	Advertising expenses	6
2	Advertising expenses to sales	34
3	R&D expenses	10
4	R&D expenses to sales	39
5	R&D expenses to assets	2
6	R&D expenses to sales, general and administrative expenses	1
7	R&D and Advertising expenses	3
8	R&D and Advertising expenses to sales	7
9	SGA expenses to sales	1
10	Advertising expenses to SGA expenses	2
11	ROA	1
12	Average of R&D intensity, advertising intensity and SGA intensity	1
13	Standardized R&D and SGA expenses	1
14	Lagged value of log of Japanese employees	1
15	Log of R&D and advertising expenses to SGA expenses	1
16	Log of R&D expenses plus one	2
17	Log of one plus Japanese employee in 1992	4
18	Log of one-year lagged R&D expenses	1
19	Log of the mean value of the ratio of R&D and advertising expense to sales over time	1
20	The mean of the ratio of advertising and R&D expenses to total assets in the previous year	1
21	The log of ownership of Japanese owners	1

Table S3-3: IV choice and frequency for research question 3

No.	IV Choices	Frequency
1	Absolute value of change in political constraint over the previous year	1
2	Equity restriction	2
3	Euro risk	7
4	Mean-centered political constraint	5
5	One hundred minus euro risk	4
6	One minus political constraint	16
7	Political constraint	34
8	Political constraint divided by Euro risk	2
9	Political dummy	2
10	Ten minus equity restriction	8

Table S3-4: Interaction choice and frequency for research question 4

No.	Interaction Choices	Frequency
1	Interaction of R&D intensity and political constraint	37
2	Interaction of advertising intensity and political constraint	29
3	Interaction of R&D intensity and euro risk	10
4	Interaction of advertising intensity and euro risk	17
5	Interaction of R&D expenses and equity restrictions	2
6	Interaction of advertising expenses and equity restrictions	2
7	Interaction of R&D expenses and political constraint	13
8	Interaction of advertising expenses and political constraint	7
9	Interaction of R&D expenses and euro risk	2
10	Interaction of R&D intensity and equity restriction	5
11	Interaction of advertising intensity and equity restriction	1
12	Interaction of R&D and advertising intensity and political constraint	15
13	Interaction of return on assets and political constraint	1

Table S3-5: DV choice and frequency for research question 1

No.	DV Choices	Frequency
1	Employee: the log of one plus the number of employees	3
2	Productivity: sales per employee	14
3	Profitability: 1=underperforming; 2= breakeven; 3=profitable	58
4	Profitability: ROA	4
5	Profitability: whether lose money	25
6	Sales: log of sales	16
7	Survival: exit or not	6

Table S3-6: DV choice and frequency for research questions 2, 3 and 4

No.	DV Choices	Freq RQ2	Freq RQ3	Freq RQ4
1	Average of ownership of largest Japanese owner over time	7	5	3
2	Ownership of largest Japanese owner	73	47	96
3	Aum of several ownerships by Japanese firms	21	17	23
4	Mean-centered Ownership of eight largest Japanese owners	4	2	8
5	Standardized ownership of largest Japanese owner	1	1	0
6	The log of ownership of Japanese owners	2	2	1
7	Whether sole ownership (1) or shared ownership (0)	4	2	2
8	Entry mode: WOS (1) or JV (0)	8	5	8

Supplement 4: Results selecting experts and specific IVs and DVs

We selected only researchers high in expertise and conducted the following analyses regarding the overall dispersion in estimates.

Table S4-1: Results from topic experts for research questions 1-4

Research Question	Positive ES & $p < 0.05$	Positive ES & $p > 0.05$	Negative ES & $p > 0.05$	Negative ES & $p < 0.05$
RQ1	26.77% (n=53)	28.28% (n=56)	26.26% (n=52)	18.69% (n=37)
RQ2	71.15% (n=74)	5.77% (n=6)	8.65% (n=9)	14.42% (n=15)
RQ3	29.89% (n=26)	11.49% (n=10)	28.74% (n=25)	29.89% (n=26)
RQ4	22.50% (n=27)	40.83% (n=49)	27.50% (n=33)	9.17% (n=11)

Notes: Topic experts are analysts from the field of strategic management and international business.

Table S4-2: Heterogeneity statistics for topic experts on strategic management and international business

Research Question	Sample size (number of analyses)	Effect Size Mean (95% CI)	Q	I ²
RQ1	198	0.048 [0.011, 0.086]	17938.19	98.90%
RQ2	104	28.966[14.816, 43.115]	2747.75	96.25%
RQ3	87	-0.325[-1.928, 1.277]	2887.62	97.02%
RQ4	120	-4.017[-14.251, 6.218]	504.84	76.43%

Notes: 47 out of 57 analysts (82.5%) from the fields of strategic management and international business.

Table S4-3: Results from analysts who do data analysis regularly

Research Question	Positive ES & $p < 0.05$	Positive ES & $p > 0.05$	Negative ES & $p > 0.05$	Negative ES & $p < 0.05$
RQ1	29.56% (n=47)	27.67% (n=44)	22.01% (n=35)	20.75% (n=33)
RQ2	70.73% (n=58)	3.66% (n=3)	7.32% (n=6)	18.29% (n=15)
RQ3	33.78% (n=25)	12.16% (n=9)	24.32% (n=18)	29.73% (n=22)
RQ4	17.82% (n=18)	37.62% (n=38)	33.66% (n=34)	10.89% (n=11)

Notes: Statistics include only the analysts who reported that they do data analysis regularly (i.e., daily, 2-3 times a week and once a week, which account for 60%, 61.2%, 67.3%, and 70.1% of the analysts for RQ1-4 respectively).

Table S4-4: Heterogeneity statistics for analysts who do data analysis regularly

Research Questions	Sample size (number of analyses)	Effect Size Mean (95% CI)	Q	I ²
RQ1	159	0.059 [0.014, 0.106]	3367.25	95.31%
RQ2	82	20.285 [9.208, 31.36]	2026.95	96%
RQ3	74	-0.219 [-1.92, 1.484]	5417.62	98.65%
RQ4	101	-8.11 [-19.83, 3.591]	373.50	73.23%

Notes: Statistics include only the analysts who reported that they do data analysis regularly (i.e., daily, 2-3 times a week and once a week, which account for 60%, 61.2%, 67.3%, and 70.1% of the analysts for RQ1-4 respectively).

Table S4-5: Results from experts on strategic management and international business who do data analysis regularly

Research Question	Positive ES & $p < 0.05$	Positive ES & $p > 0.05$	Negative ES & $p > 0.05$	Negative ES & $p < 0.05$
RQ1	28.89% (n=39)	27.41% (n=37)	25.19% (n=34)	18.52% (n=25)
RQ2	78.79% (n=52)	4.55% (n=3)	6.06% (n=4)	10.61% (n=7)
RQ3	30.30% (n=20)	13.64% (n=9)	25.76% (n=17)	30.30% (n=20)
RQ4	19.54% (n=17)	41.38% (n=36)	28.74% (n=25)	10.34% (n=9)

Notes: Statistics include only the analysts who are experts on strategic management and international business and who also do data analysis regularly (i.e., daily, 2-3 times a week and once a week, which account for 53.6%, 49.6%, 59%, 60.4% of the analysts for RQ1-4 respectively).

Table S4-6: Heterogeneity statistics for experts on strategic management and international business who do data analysis regularly

Research Questions	Sample size (number of analyses)	Effect Size Mean (95% CI)	Q	I ²
RQ1	135	0.066 [0.012, 0.12]	3241.43	95.87%
RQ2	66	28.065 [15.027, 41.102]	1693.455	96.16%
RQ3	66	-0.180 [-2.058, 1.698]	2348.138	97.23%
RQ4	87	-9.38 [-22.986, 4.219]	305.84	71.88%

Notes: Statistics include only the analysts who are experts on strategic management and international business and who also do data analysis regularly (i.e., daily, 2-3 times a week and once a week, which account for 53.6%, 49.6%, 59%, 60.4% of the analysts for RQ1-4 respectively).

Table S4-7. Exploratory analysis of patterns of IV and DV operationalizations for Research Question 2

Frequency of IV-DV pairing	DV: Operationalization of ownership by Japanese parent firm	IV: Operationalization of parent firm's intangible assets	Pattern of results	Percentage of all analyses
1st most frequent	The equity shared owned by Japanese firms	R&D expenses	Different analysts conclude effects in opposite directions	31%
2nd most frequent		The percentage of advertising expenses to sales	Analysts conclude effects in a single direction or report no effects	16%
3rd most frequent		Advertising expenses	Analysts conclude effects in a single direction or report no effects\	9%

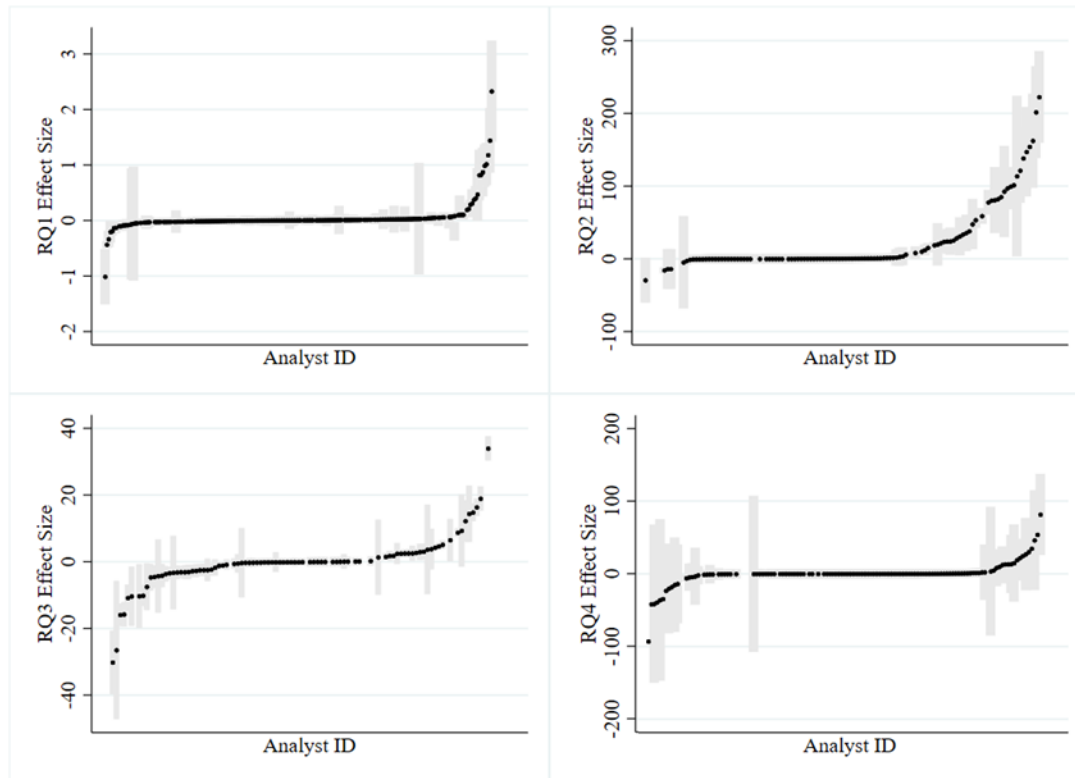


Figure S4-1: The marginal effect sizes from the four research questions, selecting topic experts on strategic management and international business.

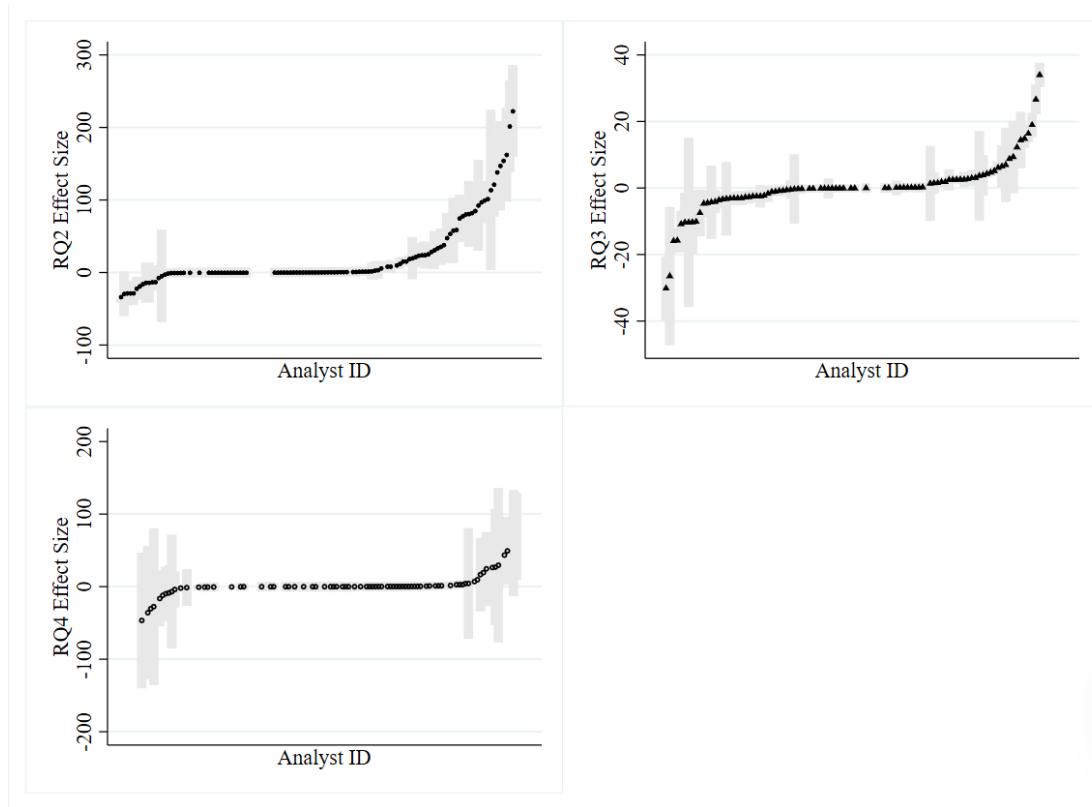


Figure S4-2: Effect size distributions for RQ2, RQ3, and RQ4 selecting specifications that used Japanese ownership as the dependent measure.

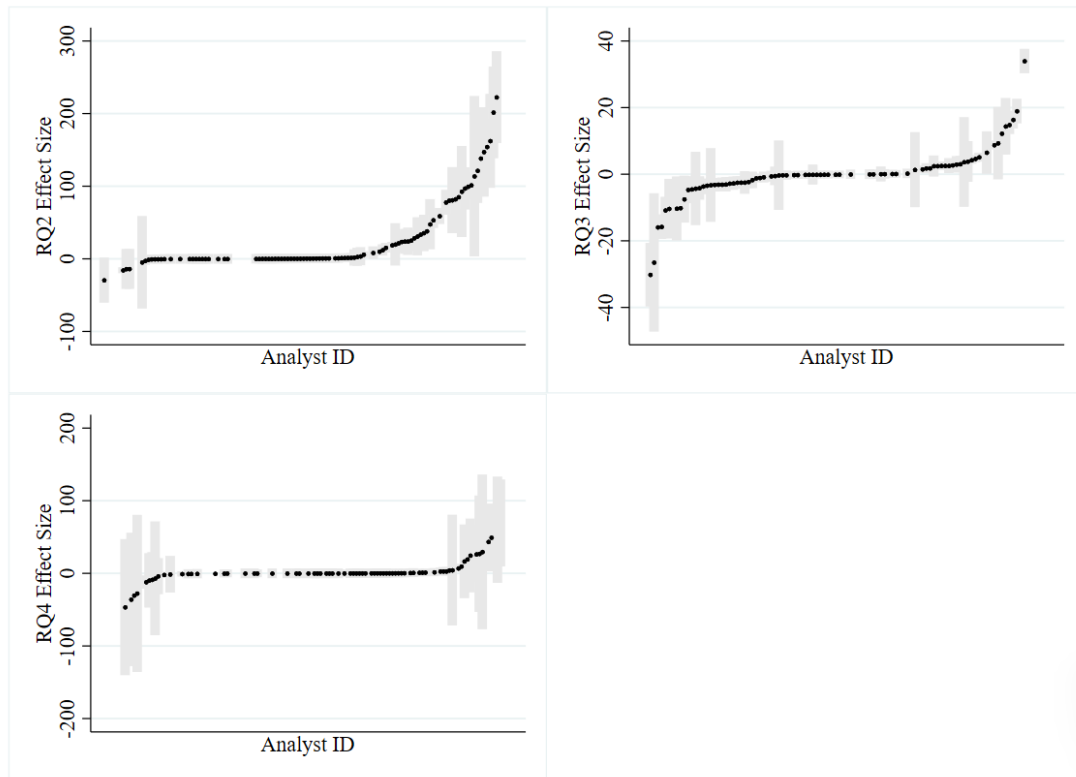


Figure S4-3: Effect size distributions for RQ2, RQ3, and RQ4 selecting expert analysts and specifications that used Japanese ownership as the dependent measure.

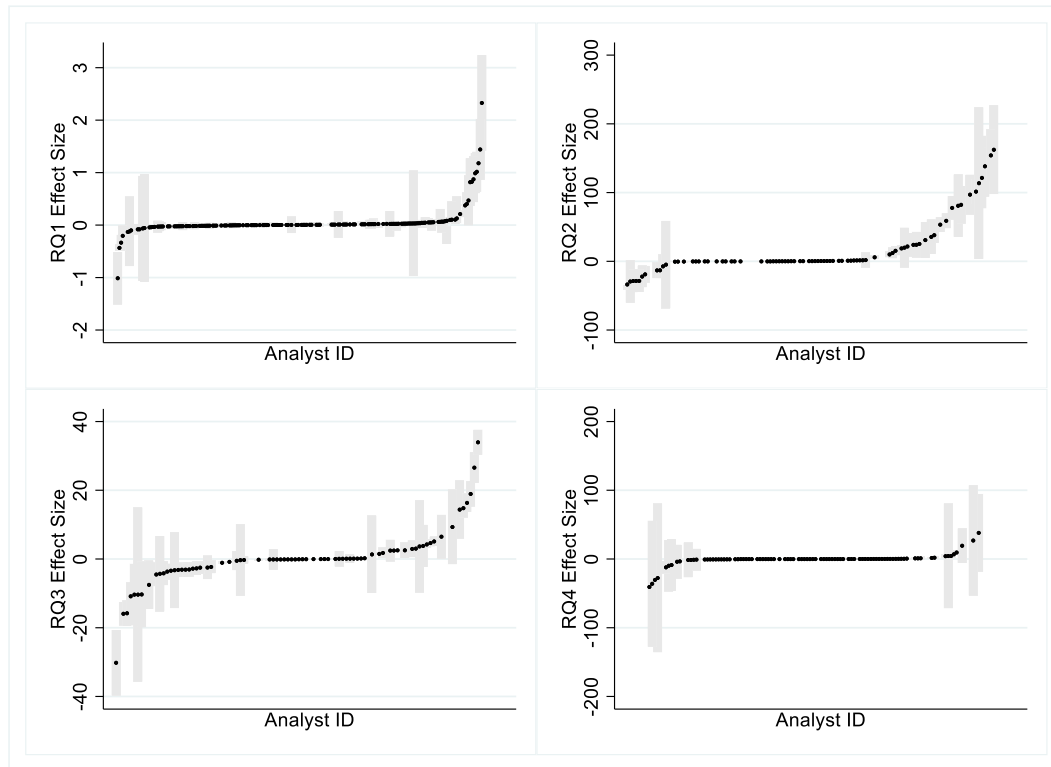


Figure S4-4: Effect size distributions selecting analysts who do data analysis regularly (i.e., daily, 2-3 times a week and once a week)

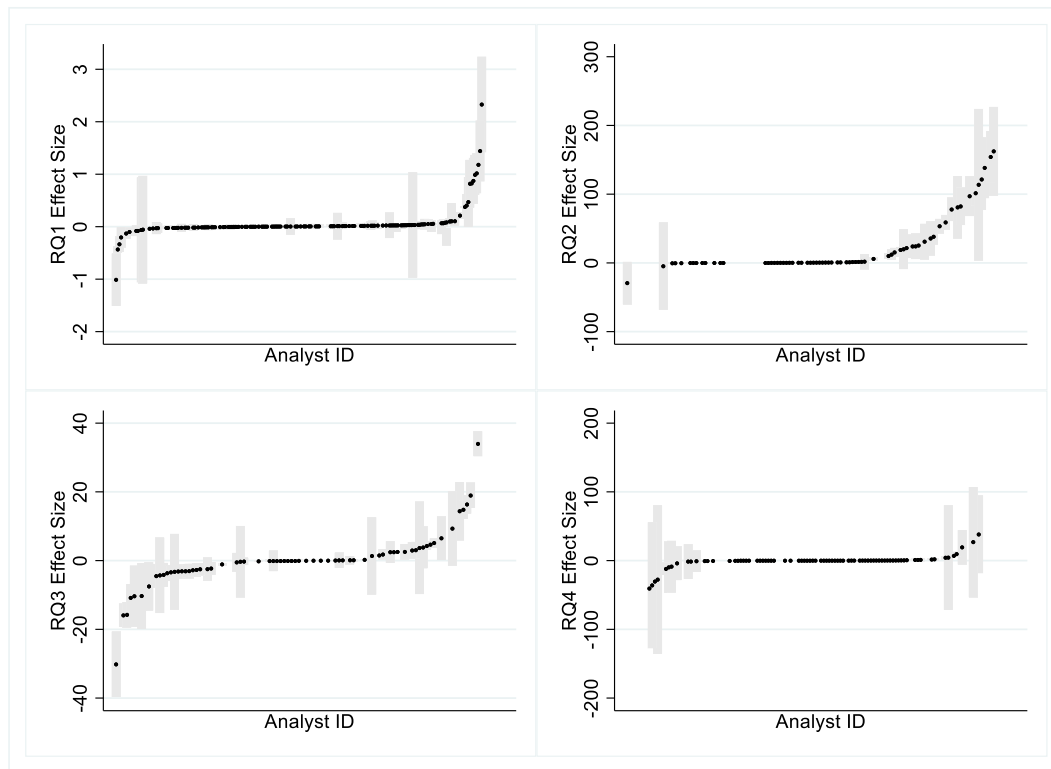


Figure S4-5: Effect size distributions selecting topic experts on strategic management and international business who also do data analysis regularly (i.e., daily, 2-3 times a week and once a week)

Supplement 5: Distribution of beliefs regarding each research question

See below for the frequency and percent of analysts endorsing different expectations regarding the research question before they received the data, and after they finished the analysis.

For RQ1, 1 = Wholly owned subsidiaries will have the highest performance, 2 = Foreign minority-owned joint ventures will have the highest performance, 3 = Foreign majority-owned joint ventures will have the highest performance, 4 = Acquisitions will have the highest performance, 5 = There will be no difference in performance by entry mode.

For research questions 2, 3 and 4, lower numbers (1) reflect the belief in a positive estimate and higher numbers (5) a belief in a negative estimate.

Table S5-1: Distribution of pre-beliefs for research question 1

RQ1	Frequency	Percent
1	28	23.53
2	14	11.76
3	30	25.21
4	8	6.72
5	39	32.77
Total	119	100

Table S5-2: Distribution of pre-beliefs for research question 2

RQ2	Frequency	Percent
1	47	39.50
2	52	43.70
3	14	11.76
4	5	4.20
5	1	0.84
Total	119	100

Table S5-3: Distribution of pre-beliefs for research question 3

RQ3	Frequency	Percent
1	5	4.20
2	9	7.56
3	16	13.45
4	50	42.02
5	39	32.77
Total	119	100

Table S5-4: Distribution of pre-beliefs for research question 4

RQ4	Frequency	Percent
1	15	12.61
2	27	22.69
3	11	9.24
4	47	39.50
5	19	15.97
Total	119	100

Table S5-5: Distribution of post-beliefs for research question 1

RQ1	Frequency	Percent
1	15	19.23
2	14	17.95
3	22	28.21
4	9	11.54
5	18	23.08
Total	78	100

Table S5-6: Distribution of post-beliefs for research question 2

RQ2	Frequency	Percent
1	27	34.62
2	37	47.44
3	6	7.69
4	6	7.69
5	2	2.56
Total	78	100

Table S5-7: Distribution of post-beliefs for research question 3

RQ3	Frequency	Percent
1	11	14.1
2	10	12.82
3	14	17.95
4	23	29.49
5	20	25.64
Total	78	100

Table S5-8: Distribution of post-beliefs for research question 4

RQ4	Frequency	Percent
1	9	11.54
2	12	15.38
3	22	28.21
4	27	34.62
5	8	10.26
Total	78	100

Table S5-9: Variance of pre-beliefs and post-survey beliefs and estimates for research questions 1 to 4

RQ	Variance of pre-beliefs	Variance of post-beliefs	Variance of estimates
1	1.495583	1.570551	.0704077
2	1.153363	1.20751	3533.138
3	1.364569	1.561568	60.51469
4	1.500257	1.504138	2351.514

Notes: For the variance of pre-survey and post-survey beliefs, as the choices are categorical variables, we use an entropy index to calculate the dispersion of survey choices.

Table S5-10: Relationships between pre-beliefs and post- beliefs for research questions 1 to 4

	Model 1	Model 2	Model 3	Model 4
Variables	Post belief on RQ1	Post belief on RQ2	Post belief on RQ3	Post belief on RQ4
Pre belief on RQ1	-0.01 [0.892]			
Pre belief on RQ2		0.21 [0.194]		
Pre belief on RQ3			0.23 [0.055]	
Pre belief on RQ4				0.10 [0.270]
Observations	78	78	78	78
Pseudo R2	0.0001	0.0089	0.0153	0.0052

Notes: Ordered probit regressions are used (p -values are in brackets).

Table S5-11: Relationships between effect size estimates and post-beliefs for research questions 1 to 4

	Model 1	Model 2	Model 3	Model 4
Variables	Post belief on RQ1	Post belief on RQ2	Post belief on RQ3	Post belief on RQ4
Effect size of estimates	-0.10 [0.888]	0.01 [0.024]	0.07 [0.009]	0.00 [0.276]
Observations	227	121	97	135
Pseudo R2	0.0001	0.0282	0.0254	0.0033

Notes: Ordered probit regressions are used (p -values are in brackets).

Supplement 6: Beliefs and variance in effect size estimates

We use ANOVA to examine whether beliefs regarding each research question before seeing the data (pre-beliefs) explain variance in estimates. The results are summarized in Tables S6-1 to S6-4.

Table S6-1: ANOVA results (DV=effect size estimates, IV=pre-beliefs regarding RQ1)

Variables	Partial SS	df	MS	F	Prob>F
Model	0.017	2	0.008	0.13	0.8778
RQ1	0.017	2	0.008	0.13	0.8778
Residual	14.31	224	0.064		
Total	14.33	224	.063		
N=227	Root MSE=0.25	R-squared=0.0012	Adjusted R-squared=-0.008		

Table S6-2: ANOVA results (DV=effect size estimates, IV=pre-beliefs regarding RQ2)

Variables	Partial SS	df	MS	F	Prob>F
Model	51669.48	3	17223.16	4.01	0.0092
RQ2	51669.48	3	17223.16	4.01	0.0092
Residual	523822.50	122	4293.63		
Total	575491.99	125	4603.94		
N=126	Root MSE=65.53	R-squared=0.090	Adjusted R-squared=0.067		

Table S6-3: ANOVA results (DV=effect size estimates, IV=pre-beliefs regarding RQ3)

Variables	Partial SS	df	MS	F	Prob>F
Model	164.15	4	41.04	0.78	0.5406
RQ3	164.15	4	41.04	0.78	0.5406
Residual	4889.28	93	52.57		
Total	5053.43	97	52.10		
N=97	Root MSE= 7.25	R-squared=0.033	Adjusted R-squared=-0.009		

Table S6-4: ANOVA results (DV=effect size estimates, IV=pre-beliefs regarding RQ4)

Variables	Partial SS	df	MS	F	Prob>F
Model	5395.19	4	1348.80	0.47	0.7583
RQ4	5395.19	4	1348.80	0.47	0.7583
Residual	37632.85	131	2875.06		
Total	382020.04	135	2875.06		
N=136	Root MSE= 53.62	R-squared=0.014	Adjusted R-squared=-0.016		

We use ANOVA to examine whether beliefs regarding each research question after seeing the results (“post-beliefs”) explain variance in estimates. The results are summarized in Tables S6-5 to S6-8.

Table S6-5: ANOVA results (DV=effect size estimates, IV=post-beliefs regarding RQ1)

Variables	Partial SS	df	MS	F	Prob>F
Model	0.106	2	0.053	0.83	0.4358
RQ1	0.106	2	0.053	0.83	0.4358
Residual	14.22	224	0.063		
Total	14.33	226	.063		
N=227	Root MSE=0.25	R-squared=0.007	Adjusted R-squared=-0.0015		

Table S6-6: ANOVA results (DV=effect size estimates, IV=post-beliefs regarding RQ2)

Variables	Partial SS	df	MS	F	Prob>F
Model	33971.68	3	11323.89	2.47	0.0651
RQ2	33971.68	3	11323.89	2.47	0.0651
Residual	53773.56	117	4579.26		
Total	569745.23	120	4747.88		
N=121	Root MSE=67.67	R-squared=0.060	Adjusted R-squared=0.036		

Table S6-7: ANOVA results (DV=effect size estimates, IV=post-beliefs regarding RQ3)

Variables	Partial SS	df	MS	F	Prob>F
Model	487.09	4	121.77	2.45	0.0513
RQ3	487.09	4	121.77	2.45	0.0513
Residual	4566.28	92	49.63		
Total	5053.37	96	52.64		
N=97	Root MSE= 7.05	R-squared=0.096	Adjusted R-squared=0.057		

Table S6-8: ANOVA results (DV=effect size estimates, IV=post-beliefs regarding RQ4)

Variables	Partial SS	df	MS	F	Prob>F
Model	9218.21	4	2304.55	0.80	0.5249
RQ4	9218.21	4	2304.55	0.80	0.5249
Residual	372740.33	130	2867.23		
Total	381958.55	134	2850.44		
N=135	Root MSE= 53.55	R-squared=0.024	Adjusted R-squared=-0.006		

Notes: Following on Silberzahn et al. (2018), since both the magnitude and precision of estimates may vary across teams, we also correlated the lower bound of the 95% CI for the effect size estimate and responses to the subjective belief items. The correlations between pre-beliefs and results for each of the four research questions are: 0.03, CI = [-0.05, 0.11]; -0.06, CI = [-0.14, 0.02]; -0.11, CI = [-0.19, -0.03]; and -0.05, CI = [-0.13, 0.03], respectively. The correlations between post-beliefs and results for each of the four research questions are: -0.07, CI = [-0.15, 0.01]; 0.12, CI = [0.04, 0.20]; -0.11, CI = [-0.19, -0.03]; and -0.12, CI = [-0.20, -0.04], respectively.

Supplement 7: Expertise, beliefs, and evidence

Table S7-1. The potential moderating role of expertise on the relationship between pre-/post-beliefs and estimates

Variables	Model 1
Pre-beliefs	4.49 (2.10) [0.033]
Post-beliefs	1.97 (4.23) [0.642]
Post-beliefs * Related field	3.32 (4.51) [0.462]
Related field	7.90 (6.26) [0.207]
Experience	0.05 (0.78) [0.946]
Frequency	-3.29 (1.38) [0.018]
Title	1.27 (2.86) [0.657]
Teaching	1.46 (4.79) [0.760]
Publications	-1.52 (2.42) [0.529]
Publication method	-1.60 (4.59) [0.727]
Female	6.29 (3.83) [0.101]
English-speaking	3.21 (3.65) [0.379]
Sample size	-0.00 (0.00) [0.650]
Research question dummies	Included
Constant	5.37 (11.85) [0.650]
Number of observations	575
F-value	4.16
R-square	0.1126

Notes: Two-tailed tests. Standard errors are in parentheses; *p*-values are in brackets.

Supplement 8: Analysts' academic profiles and demographics

Table S8-1: Distribution of analysts' publications in strategy and international business (IB) journals

Peer-reviewed Publication in Strategy or IB (number)	Frequency	Percentage
0	20	35.09
1-2	15	26.32
3-5	10	17.54
6-8	5	8.77
More than 8	7	12.28
Total	57	100

Table S8-2: Distribution of analysts' publications in methodological or statistical areas

Method Paper Publication (number)	Frequency	Percentage
0	46	80.70
1-2	8	14.04
3-5	3	5.26
Total	57	100

Table S8-3: Distribution of analysts' experience teaching statistical courses

Statistical courses taught (number)	Frequency	Percentage
0	49	85.96
1-2	4	7.02
3-5	2	3.51
More than 5	2	3.51
Total	57	100

Table S8-4: Academic positions of the analysts in our sample over the course of the project

Title	Pre-Survey	Post-Survey	Submission	Final Observation
Senior Academics	29 (24.37)	16 (20.51)	13 (18.06)	11 (19.30)
Junior Academics	89 (74.79)	62 (79.49)	59 (81.94)	46 (80.70)
Outside of Academia	1 (0.84)	0	0	0
Total	119	78	72	57

Notes: Senior academics are defined as Full Professors and Associate Professors.

Table S8-5: Highest academic degrees of the analysts in our sample over the course of the project

Degree	Pre-Survey	Post-Survey	Submission	Final Observation
Master Degree	4 (3.36)	3 (3.85)	2 (2.78)	2 (3.51)
Master/Bachelor Degree & Pursuing PhD	40 (33.61)	27 (34.62)	27 (37.50)	20 (35.09)
PhD	75 (63.03)	48 (61.54)	43 (59.72)	35 (61.40)
Total	119	78	72	57

Notes: *The four analysts with a Master Degree in the pre-survey are Research Assistants with data analysis experience.

Table S8-6: Research areas of the analysts in our sample over the course of the project

Research Area	Pre-Survey	Post-Survey	Submission	Final Observation
Strategy & IB	84 (70.58)	63 (80.77)	57 (79.17)	47 (82.46)
Others	35 (29.42)	15 (19.23)	15 (20.83)	10 (17.54)
Total	119	78	72	57

Table S8-7: Gender composition of the analysts in our sample over the course of the project

Gender	Pre-Survey	Post-Survey	Submission	Final Observation
Female	38 (31.93)	25 (32.05)	24 (33.33)	22 (38.60)
Male	81 (68.07)	53 (67.95)	48 (66.67)	35 (61.40)
Total	119	78	72	57

Table S8-8: Age distribution of the analysts in our sample over the course of the project

Age	Pre-Survey	Post-Survey	Submission	Final Observation
Average Age	36	35.13	34.82	34.88
Total	119	78	72	57

Supplement 9: Approaches to preventing and detecting strategic and inherent elasticity in analytic procedures

Below is a longer and much more detailed version of Table 5 from the main manuscript.

Approach	Description	Aimed at Inherent or Strategic Elasticity	Prevents or Detects	Calls the finding into question if...	Strengths	Weaknesses	Examples
Test- holdout sample approach	The dataset is split into two pieces, with exploratory analyses conducted on the first portion and confirmatory analyses on the second portion.	Strategic	Prevents	Results in the confirmatory sample are inconsistent with the exploratory analyses.	Prevents the original investigator from consciously or unconsciously exploiting researcher degrees of freedom to bias the results towards a particular conclusion (i.e., prevents <i>p</i> -hacking).	The analysis done in both the exploratory and confirmatory samples could be suboptimal. The findings may not be robust to alternative analytic approaches. Only addresses directional goals on the part of the analyst, not the inherent subjectivity in data analysis. Only useful for large datasets that can be divided while still retaining statistical power. Cannot be effectively employed for moderate to small datasets.	Commonly used in research on machine learning (Orrù et al., 2020; Shrestha et al., 2021). Predicting bankruptcies among companies (Altman, 1968).

Pre-registration	Researcher specifies the analytic plan before collecting or obtaining the data (Wagenmakers et al., 2012).	Strategic	Prevents	Pre-registered analysis returns a different result from a non-preregistered analysis. However, this may remain unknown unless pre-registered and other studies are compared via meta-analysis (Lodder et al., 2019), pre-registrations are compared to published reports (Goldacre et al., 2018), or analysts are randomly assigned to pre-register or not in the context of a crowd project (Mellor et al., 2023).	Analyses are publicly and transparently registered in advance, increasing trust in results regarding contentious research topics. Prevents original investigator from consciously or unconsciously exploiting researcher degrees of freedom to bias the results towards a particular conclusion (i.e., prevents <i>p</i>-hacking). Unlike the test-holdout approach, can be used for datasets that are small to moderate in size.	Reduces efficiency, and can be subjectively experienced as frustrating and burdensome by researchers (Logg & Dorison, 2021; Sarafoglou et al., 2022). The finding may not be robust to alternative analytic approaches (e.g., other researchers might have pre-registered a different and superior approach). Only addresses directional goals on the part of the analyst, not the inherent subjectivity in data analysis. The researchers may not adhere to their pre-registration or report the deviations from their original plan (Goldacre et al., 2018). Complex datasets can require exploration in the service of understanding, making the requirement to pre-specify all analyses over-restrictive (King et al., 2021).	Registered Reports (Chambers et al., 2015). Pre-registration challenge from the Open Science Center (Mellor et al., 2023). Comparison of pre-registrations of medical trials with how the trials were reported (Goldacre et al., 2018). Lodder et al. (2019) meta-analysis comparing preregistered and non-preregistered money priming studies. Adversarial collaborations (Mellers et al., 2001) in which scholars pre-specify the predictions of their competing theories.
--------------------------------	---	------------------	-----------------	--	--	--	---

Results blind review	Reviewers assess and make publication recommendations based solely on the introduction and methods sections of the paper. The results and discussion sections are withheld from reviewers.	Strategic	Prevents	Authors who know their paper's results will not affect publication decisions are less likely to obtain empirical support for the hypothesis.	Removes the publication incentive for authors to find $p < 0.05$ support for their hypothesis, or results preferred by reviewers.	Removes information from the research report that many reviewers may consider highly relevant to contribution (i.e., the empirical results). Does not affect the researchers' intrinsic motivation to obtain results favoring theories to which they are intellectually committed. Does not address inherent subjectivity in data analysis.	Experimental work randomly assigning academic to reviewers to standard or results-blind review (Maggin et al., 2020). Highlighted for potential use by strategic management journals by Goldfarb & King (2016).
Data-blind analysis	Member of the original research team relabels or recodes some of the key variables before the analyses are conducted	Strategic	Prevents	Data-blind analysis returns different results from a non-blind analysis. However, this direct comparison is usually not made.	Prevents original investigator from consciously or unconsciously exploiting researcher degrees of freedom to bias the results towards a particular conclusion (i.e., prevents p-hacking). Does not sacrifice statistical power for confirmatory tests like the test-holdout sample approach.	Reduces efficiency in that it takes time and care to blind and then unblind data. The data-blind analysis could be suboptimal, and its robustness to alternative approaches remains unknown. Removes directional goals from the equation, but not the inherent subjectivity of data analysis.	Widely used in some areas of nuclear and particle physics (MacCoun & Perlmutter, 2015). Used in measurements of supernovas in studies of the expansion of the universe (Conley et al., 2006).

Independent re-analysis	Independent researcher uses an alternative analytic approach to test the same theoretical idea on the same data.	Both	Detects	The results of the alternative analysis are meaningfully different, and the alternative specification is more valid.	Post-publication peer review approach can highlight shortcomings of published analyses.	Incentive to disprove the original finding (overturn bias) (Galiani et al., 2017).	Simonsohn (2011a, 2011b) critiques of name-letter effects. Goldfarb & Yan's (2021) re-analysis of the categorical imperative effect.
Error analysis	Independent researcher identifies errors in data handling or analysis.	Both	Detects	The errors systematically favor the original hypothesis. Removing the error or errors eliminates the effect.	If properly identified, errors may be objectively verified by further independent researchers. Tests can be carried out showing that errors favoring the hypothesis outnumber those disfavoring the hypothesis in a manner unlikely to occur randomly.	Identification of errors may be partly subjective; critics may selectively seek or overcount errors disfavoring the original hypothesis. Either requires or is greatly facilitated by access to the original dataset(s).	Rosenthal (1978) demonstration that data errors systematically favor the original researchers' hypothesis by a ratio of 2 to 1.

Statistical tests for publication bias and questionable research practices	Meta-scientific analysis of a study set or individual study to test for omitted studies and data patterns consistent with QRPs.	Strategic	Detects	Published studies on a topic show strong signs of publication bias and exploitation of researcher degrees of freedom (e.g., many p-values that barely cross the $p < .05$ threshold).	No need to access the original dataset or formulate an alternative analytic strategy. Invaluable approach to assess the evidentiary value of a body of published literature.	Publication bias is pervasive in academic research. Thus, any sufficiently large study set will likely test positive. For critiques of single papers, the danger of committing the <i>prosecutor's fallacy</i> (i.e., assuming the probability of a random match is equal to the probability of innocence). Publication bias tests can also exhibit publication bias and be cherry-picked (Simonsohn, 2013).	Francis (2012, 2013) critiques of individual papers based on Ioannidis and Trikalinos (2007) test for excess significance. P-curve analysis of choice overload literature (Simonsohn et al., 2014). Carter & McCullough (2014) publication bias tests of ego depletion literature. Incredibility index analysis of Bem (2011) “feeling the future” ESP studies (Schimmack, 2012). “Scientific apophenia” results of Goldfarb & King (2016) regarding the strategic management literature.
--	--	------------------	----------------	---	--	---	---

“Chrysalis” analysis	Comparison of earlier vs. later versions of research report (e.g., doctoral dissertations vs. published reports of the same work) (O’Boyle et al., 2017).	Strategic	Detects	Later reports exclude measures and observations associated with $p > .05$ findings or nonsupport for key theoretical predictions in earlier reports (i.e., a “chrysalis effect”).	No need to access the original dataset or formulate an alternative analytic strategy.	Limited to research reports for which earlier versions are retrievable. Cannot detect QRPs used in the first available reports of the research.	Ratio of supported to unsupported hypotheses doubles between dissertations and final published articles in management (O’Boyle et al., 2017).
-----------------------------	---	-----------	---------	---	---	--	---

Conducting an independent replication study	Independent research team collects new data using the same methodology and runs the same analyses as described in the original paper.	Strategic	Detects	Original finding does not emerge again in the independent replication.	No need to access the original dataset or formulate an alternative analytic strategy. In crowd replication projects, can be combined with publication bias tests (e.g., effects with small p values were more likely to replicate than those with large p values, e.g., ≥ 0.05) (Open Science Collaboration, 2015).	Attributional ambiguity is high, since findings may fail to replicate for reasons other than exploitation of researcher degrees of freedom in the original paper. These include cultural, population, and linguistic differences; lack of fidelity to the original method; and random chance. P-hacked false positive findings will generally fail to replicate, but so will many findings that were not p-hacked. Labor intensive in that new data must be collected. Reverse publication bias in favor of failed replications (Galiani et al., 2017).	Reproducibility Project: Psychology (Open Science Collaboration, 2015). Many Labs initiatives (e.g. Klein et al., 2014; Klein et al., 2018). One-shot replications by individual laboratories (LeBel, 2015).
--	---	-----------	---------	--	---	--	---

Red team approach	Independent research team attempts to debias and optimize an original investigation before it is submitted for publication (Lakens, 2020).	Strategic	Prevents	Red team identifies biased or suboptimal analytic approach.	Adversarial approach can identify problematic methods and analyses before the paper is published.	Inefficient in that more researchers and time are required to complete the project. If the red team is provided authorship, they are disincentivized to find fatal problems that preclude publication. Costs money if the red team is rewarded financially rather than with authorship; less well funded laboratories may be unable to afford the paid approach.	Coles et al. (2022) red team challenge for facial feedback effects. Schaerer et al. (2023) red team challenge regarding the trajectory of gender discrimination over time.
Retrospective re-assessment	Original researcher decides after-the-fact that the analysis was sub-optimal, and a different approach would be better.	Both	Detects	The retrospectively preferred analytic approach would have produced qualitatively different results.	Original researchers have privileged knowledge of the process by which the research was done.	Unlikely to occur due to motivated reasoning and professional incentive structure discouraging scholars from repudiating their own work. Memory and introspection could be faulty. Original author's opinion only a heuristic cue to quality of the analysis.	Carney (2016) repudiation of original Power Posing analysis for exploiting researcher degrees of freedom. Loss-of-confidence (LOC) declarations from authors admitting to p-hacking and analytic mistakes (Rohrer et al., 2021). "Mega-corrections" in which original authors present new analytic approach and results (e.g. Silberzahn et al., 2014).

Comparing employed analyses	A single researcher surveys the various analytic approaches used in different studies on the same topic, and potentially applies them all to the same data.	Strategic	Detects	There are inconsistencies in analytic approaches across and within research teams that are difficult to defend. Analyses employed in other papers return less favorable results for the theoretical hypothesis than the original analysis.	Alternative analytic approaches already have endorsement of topic experts as valid and defensible.	The analyses used in a set of articles may be suboptimal relative to unused analytic approaches. Different studies may be analyzed in different ways for valid reasons the re-analyst does not understand or acknowledge.	Elson et al. (2014) alternative analyses of the competitive reaction time task, a common dependent measure in laboratory studies of aggression. Harris et al. (2013) examination of alternative cutoffs used for fertile and nonfertile phases in research on menstrual cycles and mate preferences.
Strategic analysis	A crowd of analysts is asked to either produce $p < 0.05$ support for the original hypothesis or provided no goal.	Strategic	Detects	May show that an effect is only obtained when analysts are instructed to p -hack.	Demonstrates how intentional exploitation of researcher degrees of freedom can bias the research results.	Cannot rule out “winner’s curse” in the original analysis or demonstrate the goals of the original researchers. Extremely inefficient in that a crowd of researchers must be recruited.	Ongoing crowd initiative via the University of Pennsylvania’s Adversarial Collaboration Project.

Multiverse analysis or Specification curve	One analyst employs numerous specifications (Simonsohn et al., 2020; Steegen et al., 2016).	Inherent	Detects	Finding only supported in a small minority of legitimate specifications.	Robustness of the finding to thousands, even millions of analytic approaches is examined.	Requires access to computational resources to run numerous analyses. Potentially subject to the biases of the one analyst or small team of analysts who plans the multiverse.	Sala-i-Martin's (1997) two billion regressions examining predictors of economic growth. re-analysis of research on fertility cycles. Orben & Przybylski (2019) re-analysis of screen time and well-being. Smerdon et al. (2020) re-analysis of Stafford (2018) dataset on stereotype threat in female chess players.
Many analysts	Many researchers analyze the same data to test the same hypothesis (Silberzahn et al., 2018).	Inherent	Detects	Original finding is not supported in numerous independent re-analyses conducted by qualified experts.	Not subject to the intellectual allegiances (Berman & Reich, 2010) of any individual researcher. Lack of perverse publication incentives to achieve $p < 0.05$ or disprove the original finding.	Unlikely to capture all defensible specifications. Extremely inefficient in that a crowd of researchers must be recruited. Less expert members of the crowd may carry out suboptimal specifications.	Silberzahn et al. (2018) many analysts approach to referee bias based on player race. Schweinsberg et al. (2021) many analysts approach to gender and status in scientific debates. Botvinik-Nezer et al. (2020) many analysts approach to fMRI dataset. Menkveld et al. (2024) many analysts initiative in finance.

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589–609. <https://doi.org/10.2307/2978933>
- Berman, J. S., & Reich, C. M. (2010). Investigator allegiance and the evaluation of psychotherapy outcome research. *European Journal of Psychotherapy & Counselling*, 12(1), 11–21. <https://doi.org/10.1080/13642531003637775>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Benoit, R. G., Berkers, R. M. W. J., Bhanji, J. P., Biswal, B. B., Bobadilla-Suarez, S., Bortolini, T., Bottenhorn, K. L., Bowring, A., Braem, S., Brooks, H. R., Brudner, E. G., Calderon, C. B., Camilleri, J. A., Castrellon, J. J., Cecchetti, L., Cieslik, E. C., Cole, Z. J., Collignon, O., Cox, R. W., Cunningham, W. A., Czoschke, S., Dadi, K., Davis, C. P., Luca, A. D., Delgado, M. R., Demetriou, L., Dennison, J. B., Di, X., Dickie, E. W., Dobryakova, E., Donnat, C. L., Dukart, J., Duncan, N. W., Durnez, J., Eed, A., Eickhoff, S. B., Erhart, A., Fontanesi, L., Fricke, G. M., Fu, S., Galván, A., Gau, R., Genon, S., Glatard, T., Glerean, E., Goeman, J. J., Golowin, S. A. E., González-García, C., Gorgolewski, K. J., Grady, C. L., Green, M. A., Moreira, J. F. G., Guest, O., Hakimi, S., Hamilton, J. P., Hancock, R., Handjaras, G., Harry, B. B., Hawco, C., Herholz, P., Herman, G., Heunis, S., Hoffstaedter, F., Hogeveen, J., Holmes, S., Hu, C.-P., Huettel, S. A., Hughes, M. E., Iacovella, V., Iordan, A. D., Isager, P. M., Isik, A. I., Jahn, A., Johnson, M. R., Johnstone, T., Joseph, M. J. E., Juliano, A. C., Kable, J. W., Kassinopoulos, M., Koba, C., Kong, X.-Z., Kosciuk, T. R., Kucukboyaci, N. E., Kuhl, B. A., Kupek, S., Laird, A. R., Lamm, C., Langner, R., Lauharatanahirun, N., Lee, H., Lee, S., Leemans, A., Leo, A., Lesage, E., Li, F., Li, M. Y. C., Lim, P. C., Lintz, E. N., Liphardt, S. W., Losecaat Vermeer, A. B., Love, B. C., Mack, M. L., Malpica, N., Marins, T., Maumet, C., McDonald, K., McGuire, J. T., Melero, H., Méndez Leal, A. S., Meyer, B., Meyer, K. N., Mihai, G., Mitsis, G. D., Moll, J., Nielson, D. M., Nilsson, G., Notter, M. P., Olivetti, E., Onicas, A. I., Papale, P., Patil, K. R., Peelle, J. E., Pérez, A., Pischella, D., Poline, J.-B., Prystauka, Y., Ray, S., Reuter-Lorenz, P. A., Reynolds, R. C., Ricciardi, E., Rieck, J. R., Rodriguez-Thompson, A. M., Romy, A., Salo, T., Samanez-Larkin, G. R., Sanz-Morales, E., Schlichting, M. L., Schultz, D. H., Shen, Q., Sheridan, M. A., Silvers, J. A., Skagerlund, K., Smith, A., Smith, D. V., Sokol-Hessner, P., Steinkamp, S. R., Tashjian, S. M., Thirion, B., Thorp, J. N., Tinghög, G., Tisdall, L., Thompson, S. H., Toro-Serey, C., Torre Tresols, J. J., Tozzi, L., Truong, V., Turella, L., van 't Veer, A. E., Verguts, T., Vettel, J. M., Vijayarajah, S., Vo, K., Wall, M. B., Weeda, W. D., Weis, S., White, D. J., Wisniewski, D., Xifra-Porxas, A., Yearling, E. A., Yoon, S., Yuan, R., Yuen, K. S. L., Zhang, L., Zhang, X., Zosky, J. E., Nichols, T. E., Poldrack, R. A., & Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Carney, D. R. (2016). *My position on “Power Poses”*. [Unpublished manuscript]. University of California, Berkeley. Retrieved from https://faculty.haas.berkeley.edu/dana_carney/pdf_my%20position%20on%20power%20poses.pdf

- Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5, 823. <https://doi.org/10.3389/fpsyg.2014.00823>
- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex*, 66, A1–2. <https://doi.org/10.1016/j.cortex.2015.03.022>
- Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L. G., Willis, M. L., Foroni, F., Reggev, N., Mokady, A., Forscher, P. S., Hunter, J. F., Kaminski, G., Yüvrük, E., Kapucu, A., Nagy, T., Hajdu, N., Tejada, J., Freitag, R. M. K., Zambrano, D., Som, B., Aczel, B., Barzykowski, K., Adamus, S., Filip, K., Yamada, Y., Ikeda, A., Eaves, D. L., Levitan, C. A., Leiweke, S., Parzuchowski, M., Butcher, N., Pfuhl, G., Basnight-Brown, D. M., Hinojosa, J. A., Montoro, P. R., Javela D, L. G., Vezirian, K., Ijzerman, H., Trujillo, N., Pressman, S. D., Gygas, P. M., Özdoğru, A. A., Ruiz-Fernandez, S., Ellsworth, P. C., Gaertner, L., Strack, F., Marozzi, M., & Liuzza, M. T. (2022). A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration. *Nature Human Behaviour*, 6(12), 1731-1742. <https://doi.org/10.1038/s41562-022-01458-9>
- Conley, A., Goldhaber, G., Wang, L., Aldering, G., Amanullah, R., Commins, E. D., Fadeyev, V., Folatelli, G., Garavini, G., Gibbons, R., Goobar, A., Groom, D. E., Hook, I., Howell, D. A., Kim, A. G., Knop, R. A., Kowalski, M., Kuznetsova, N., Lidman, C., Nobili, S., Nugent, P. E., Pain, R., Perlmutter, S., Smith, E., Spadafora, A. L., Stanishev, V., Strovink, M., Thomas, R. C., & Wood-Vasey, W. M. (2006). Measurement of Ω_m , Ω_Λ from a blind analysis of type Ia supernovae with CMAGIC: Using color information to verify the acceleration of the universe. *The Astrophysical Journal*, 644(1), 1-20. <https://doi.org/10.1086/503533>
- Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment*, 26(2), 419-432. <https://doi.org/10.1037/a0035569>
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19(2), 151-156. <https://doi.org/10.3758/s13423-012-0227-9>
- Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, 57(5), 153-169. <https://doi.org/10.1016/j.jmp.2013.02.003>
- Galiani, S., Gertler, P., & Romero, M. (2017). Incentives for replication in economics. NBER working paper. Available at: https://www.nber.org/system/files/working_papers/w23576/w23576.pdf
- Goldacre, B., DeVito, N. J., Heneghan, C., Irving, F., Bacon, S., Fleminger, J., & Curtis, H. (2018). Compliance with requirement to report results on the EU Clinical Trials Register: Cohort study and web resource. *British Medical Journal*, 362, k3218. <https://doi.org/10.1136/bmj.k3218>
- Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, 37(1), 167-176. <https://doi.org/10.1002/smj.2459>

- Goldfarb, B., & Yan, L. (2021). Revisiting Zuckerman's (1999) categorical imperative: An application of epistemic maps for replication. *Strategic Management Journal*, 42(11), 1963-1992. <https://doi.org/10.1002/smj.3290>
- Harris, C. R., Chabot, A., & Mickes, L. (2013). Shifts in methodology and theory in menstrual cycle research. *Sex Roles*, 69(9), 525–535. <https://doi.org/10.1007/s11199-013-0302-3>
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Journal of Clinical Trials*, 4(3), 245–253. <https://doi.org/10.1177/1740774507079441>
- King, A., Goldfarb, B., & Simcoe, T. (2021). Learning from testimony on quantitative research in management. *Academy of Management Review*, 46(3), 465–488. <https://doi.org/10.5465/amr.2018.0421>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., Hasselman, F., Hicks, J. A., Hovermale, J. F., Hunt, S. J., Huntsinger, J. R., Ijzerman, H., John, M.-S., Joy-Gaba, J. A., Kappes, H. B., Krueger, L. E., Kurtz, J., Levitan, C. A., Mallett, R. K., Morris, W. L., Nelson, A. J., Nier, J. A., Packard, G., Pilati, R., Rutchick, A. M., Schmidt, K., Skorinko, J. L., Smith, R., Steiner, T. G., Storbeck, J., van Swol, L. M., Thompson, D., van 't Veer, A. E., Vaughn, L. A., Vranka, M., Wichman, A. L., Woodzicka, J. A., & Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45(3), 142-152. <https://doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., Rédei, A. C., Cai, H., Cambier, F., Cantarero, K., Carmichael, C. L., Ceric, F., Chandler, J., Chang, J.-H., Chatard, A., Chen, E. E., Cheong, W., Cicero, D. C., Coen, S., Coleman, J. A., Collisson, B., Conway, M. A., Corker, K. S., Curran, P. G., Cushman, F., Dagona, Z. K., Dalgard, I., Dalla Rosa, A., Davis, W. E., de Bruijn, M., de Schutter, L., Devos, T., de Vries, M., Doğulu, C., Dozo, N., Dukes, K. N., Dunham, Y., Durrheim, K., Ebersole, C. R., Edlund, J. E., Eller, A., English, A. S., Finck, C., Frankowska, N., Freyre, M.-Á., Friedman, M., Galliani, E. M., Gandhi, J. C., Ghoshal, T., Giessner, S. R., Gill, T., Gnamb, T., Gómez, Á., González, R., Graham, J., Grahe, J. E., Grahek, I., Green, E. G. T., Hai, K., Haigh, M., Haines, E. L., Hall, M. P., Heffernan, M. E., Hicks, J. A., Houdek, P., Huntsinger, J. R., Huynh, H. P., Ijzerman, H., Inbar, Y., Innes-Ker, Å. H., Jiménez-Leal, W., John, M.-S., Joy-Gaba, J. A., Kamiloğlu, R. G., Kappes, H. B., Karabati, S., Karick, H., Keller, V. N., Kende, A., Kervyn, N., Knežević, G., Kovacs, C., Krueger, L. E., Kurapov, G., Kurtz, J., Lakens, D., Lazarević, L. B., Levitan, C. A., Lewis, N. A., Lins, S., Lipsey, N. P., Losee, J. E., Maassen, E., Maitner, A. T., Malingumu, W., Mallett, R. K., Marotta, S. A., Mededović, J., Mena-Pacheco, F., Milfont, T. L., Morris, W. L., Murphy, S. C., Myachikov, A., Neave, N., Neijenhuijs, K., Nelson, A. J., Neto, F., Nichols, A. L., Ocampo, A., O'Donnell, S. L., Oikawa, H., Oikawa, M., Ong, E., Orosz, G., Osowiecka, M., Packard, G., Pérez-Sánchez, R., Petrović, B., Pilati, R., Pinter, B., Podesta, L., Pogge, G., Pollmann, M. M. H., Rutchick, A. M., Saavedra, P., Saeri, A. K., Salomon, E., Schmidt, K., Schönbrodt, F. D., Sekerdej, M. B., Sirlopú, D., Skorinko, J. L. M., Smith, M. A., Smith-Castro, V., Smolders, K. C. H. J., Sobkow, A.,

- Sowden, W., Spachtholz, P., Srivastava, M., Steiner, T. G., Stouten, J., Street, C. N. H., Sundfelt, O. K., Szeto, S., Szumowska, E., Tang, A. C. W., Tanzer, N., Tear, M. J., Theriault, J., Thomae, M., Torres, D., Traczyk, J., Tybur, J. M., Ujhelyi, A., van Aert, R. C. M., van Assen, M. A. L. M., van der Hulst, M., van Lange, P. A. M., van 't Veer, A. E., Vázquez- Echeverría, A., Vaughn, L. A., Vázquez, A., Vega, L. D., Verniers, C., Verschoor, M., Voermans, I. P. J., Vranka, M. A., Welch, C., Wichman, A. L., Williams, L. A., Wood, M., Woodzicka, J. A., Wronska, M. K., Young, L., Zelenski, J. M., Zhijia, Z., & Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490. <https://doi.org/10.1177/2515245918810225>
- Lakens, D. (2020). Pandemic researchers—recruit your own best critics. *Nature*, 581(7807), 121–122. <https://doi.org/10.1038/d41586-020-01392-8>.
- Landy, J. F., Jia, M., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M., Derks, K., Wagenmakers, E.-J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., Cheung, F., Cimpian, A., Dohle, S., Donnellan, M. B., Hahn, A., Hall, M. P., Jiménez-Leal, W., Johnson, D. J., Lucas, R. E., Monin, B., Montealegre, A., Mullen, E., Pang, J., Ray, J., Reinero, D. A., Reynolds, J., Sowden, W., Storage, D., Su, R., Tworek, C. M., van Bavel, J. J., Walco, D., Wills, J., Xu, X., Yam, K. C., Yang, X., Cunningham, W. A., Schweinsberg, M., Urwitz, M., & Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, 146(5), 451-479. <https://doi.org/10.1037/bul0000220>
- LeBel, E. P. (2015). A new replication norm for psychology. *Collabra*, 1(1), 1–13. <https://doi.org/10.1525/collabra.23>
- Lodder, P., Ong, H. H., Grasman, R. P., & Wicherts, J. M. (2019). A comprehensive meta-analysis of money priming. *Journal of Experimental Psychology: General*, 148(4), 688–712. <https://doi.org/10.1037/xge0000570>
- Logg, J. M., & Dorison, C. A. (2021). Pre-registration: Weighing costs and benefits for researchers. *Organizational Behavior and Human Decision Processes*, 167, 18–27. <https://doi.org/10.1016/j.obhdp.2021.05.006>
- MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, 526(7572), 187–189. <https://doi.org/10.1038/526187a>
- Maggin, D. M., Robertson, R. E., & Cook, B. G. (2020). Introduction to the special series on results-blind peer review: An experimental analysis on editorial recommendations and manuscript evaluations. *Behavioral Disorders*, 45(4), 195-206. <https://doi.org/10.1177/0198742920936619>
- Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction Effects? An exercise in adversarial collaboration. *Psychological Science*, 12(4), 269-275. <https://doi.org/10.1111/1467-9280.00350>
- Mellor, D. T., Esposito, J., Hardwicke, T. E., Nosek, B. A., Cohoon, J., Soderberg, C. K., Kidwell, M. C., Clyburne-Sherin, A., Buck, S., DeHaven, A. C., & Speidel, R. (2023). *Preregistration challenge: Plan, test, discover*. Retrieved from <https://osf.io/x5w7h>.
- Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Neusüß, S., Razen, M., Weitzel, U., Abad-Díaz, D., Abudy, M., Adrian, T., Ait-Sahalia, Y., Akmansoy, O., Alcock, J. T.,

Alexeev, V., Aloosh, A., Amato, L., Amaya, D., Angel, J. J., Avetikian, A. T., Bach, A., Baidoo, E., Bakalli, G., Bao, L., Barbon, A., Bashchenko, O., Bindra, P. C., Bjønnes, G. H., Black, J. R., Black, B. S., Bogoev, D., Correa, S. B., Bondarenko, O., Bos, C. S., Bosch-Rosa, C., Bourri, E., Brownlees, C., Calamia, A., Cao, V. N., Capelle-Blancard, G., Romero, L. M. C., Caporin, M., Carrion, A., Caskurlu, T., Chakrabarty, B., Chen, J., Chernov, M., Cheung, W., Chincarini, L. B., Chordia, T., Chow, S.-C., Clapham, B., Colliard, J.-E., Comerton-Forde, C., Curran, E., Dao, T., Dare, W., Davies, R. J., Blasis, R. D., Nard, G. F. D., Declerck, F., Deev, O., Degryse, H., Deku, S. Y., Desagre, C., Dijk, M. A. V., Dim, C., Dimpfl, T., Dong, Y. J., Drummond, P. A., Dudda, T., Duevski, T., Dumitrescu, A., Dyakov, T., Dyhrberg, A. H., Dzieliński, M., Eksi, A., Kalak, I. E., Ellen, S. T., Eugster, N., Evans, M. D. D., Farrell, M., Felez-Vinas, E., Ferrara, G., Ferrouhi, E. M., Flori, A., Fluharty-Jaidee, J. T., Foley, S. D. V., Fong, K. Y. L., Foucault, T., Franus, T., Franzoni, F., Frijns, B., Frömmel, M., Fu, S. M., Füllbrunn, S. C., Gan, B., Gao, G., Gehrig, T. P., Gemayel, R., Gerritsen, D., Gil-Bazo, J., Gilder, D., Glosten, L. R., Gomez, T., Gorbenko, A., Grammig, J., Grégoire, V., Güçbilmez, U., Hagströmer, B., Hambuckers, J., Hapnes, E., Harris, J. H., Harris, L., Hartmann, S., Hasse, J.-B., Hautsch, N., He, X.-Z., Heath, D., Hediger, S., Hendershott, T., Hibbert, A. M., Hjalmarsson, E., Hoelscher, S. A., Hoffmann, P., Holden, C. W., Horenstein, A. R., Huang, W., Huang, D., Hurlin, C., Ilczuk, K., Ivashchenko, A., Iyer, S. R., Jahanshahloo, H., Jalkh, N., Jones, C. M., Jurkatis, S., Jylhä, P., Kaeck, A. T., Kaiser, G., Karam, A., Karmaziene, E., Kassner, B., Kaustia, M., Kazak, E., Kearney, F., Kervel, V. V., Khan, S. A., Khomyn, M. K., Klein, T., Klein, O., Klos, A., Koetter, M., Kolokolov, A., Korajczyk, R. A., Kozhan, R., Krahnen, J. P., Kuhle, P., Kwan, A., Lajaunie, Q., Lam, F. Y. E. C., Lambert, M., Langlois, H., Lausen, J., Lauter, T., Leippold, M., Levin, V., Li, Y., Li, H., Liew, C. Y., Lindner, T., Linton, O., Liu, J., Liu, A., Llorente, G., Lof, M., Lohr, A., Longstaff, F., Lopez-Lira, A., Mankad, S., Mano, N., Marchal, A., Martineau, C., Mazzola, F., Meloso, D., Mi, M. G., Mihet, R., Mohan, V., Moinas, S., Moore, D., Mu, L., Muravyev, D., Murphy, D., Neszveda, G., Neumeier, C., Nielsson, U., Nimalendran, M., Nolte, S., Norden, L. L., O'Neill, P., Obaid, K., Ødegaard, B. A., Östberg, P., Pagnotta, E., Painter, M., Palan, S., Palit, I. J., Park, A., Pascual, R., Pasquariello, P., Pastor, L., Patel, V., Patton, A. J., Pearson, N. D., Pelizzon, L., Pelli, M., Pelster, M., Pérignon, C., Pfiffer, C., Philip, R., Plíhal, T., Prakash, P., Press, O.-A., Prodromou, T., Prokopczuk, M., Putnins, T., Qian, Y., Raizada, G., Rakowski, D., Ranaldo, A., Regis, L., Reitz, S., Renault, T., Renjie, R. W., Reno, R., Riddiough, S. J., Rinne, K., Rintamäki, P., Riordan, R., Rittmannsberger, T., Longarela, I. R., Roesch, D., Rognone, L., Roseman, B., Roşu, I., Roy, S., Rudolf, N., Rush, S. R., Rzayev, K., Rzeźnik, A. A., Sanford, A., Sankaran, H., Sarkar, A., Sarno, L., Scaillet, O., Scharnowski, S., Schenk-Hoppé, K. R., Schertler, A., Schneider, M., Schroeder, F., Schürhoff, N., Schuster, P., Schwarz, M. A., Seasholes, M. S., Seeger, N. J., Shachar, O., Shkilko, A., Shui, J., Sikic, M., Simion, G., Smales, L. A., Söderlind, P., Sojli, E., Sokolov, K., Sönksen, J., Spokeviciute, L., Stefanova, D., Subrahmanyam, M. G., Szaszi, B., Talavera, O., Tang, Y., Taylor, N., Tham, W. W., Theissen, E., Thimme, J., Tonks, I., Tran, H., Trapin, L., Trolle, A. B., Vaduva, M. A., Valente, G., Ness, R. A. V., Vasquez, A., Verousis, T., Verwijmeren, P., Vilhelmsson, A., Vilkov, G., Vladimirov, V., Vogel, S., Voigt, S., Wagner, W., Walther, T., Weiss, P., Wel, M. V. D., Werner, I. M., Westerholm, P. J., Westheide, C., Wika, H. C., Wipplinger, E., Wolf, M., Wolff, C. C. P., Wolk, L., Wong, W.-K., Wrampelmeyer, J., Wu, Z.-X., Xia, S., Xiu, D., Xu, K., Xu, C., Yadav, P. K., Yagüe, J.,

- Yan, C., Yang, A., Yoo, W., Yu, W., Yu, Y., Yu, S., Yueshen, B. Z., Yuferova, D., Zamojski, M., Zareei, A., Zeisberger, S. M., Zhang, L., Zhang, S. S., Zhang, X., Zhao, L., Zhong, Z., Zhou, Z. I., Zhou, C., Zhu, X. S., Zoican, M., & Zwinkels, R. (2024). Nonstandard Errors. *The Journal of Finance*, 79(3), 2339-2390. <https://doi.org/10.1111/jofi.13337>
- O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 43(2), 376-399.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, 3(2), 173–182. <https://doi.org/10.1038/s41562-018-0506-1>
- Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology*, 10, 1–10. <https://doi.org/10.3389/fpsyg.2019.02970>
- Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B., Schmukle, S. C., Silberzahn, R., Willén, R. M., Carlsson, R., Lucas, R. E., Strand, J., Vazire, S., Witt, J. K., Zentall, T. R., Chabris, C. F., & Yarkoni, T. (2021). Putting the self in self-correction: Findings from the loss-of-confidence project. *Perspectives on Psychological Science*, 16(6), 1255-1269. <https://doi.org/10.1177/1745691620964106>
- Rosenthal, R. (1978). How often are our numbers wrong? *American Psychologist*, 33, 1005–1008. <https://doi.org/10.1037/0003-066X.33.11.1005>
- Sala-i-Martin, X. (1997). I just ran two million regressions. *American Economic Review*, 87(2), 178-183. <http://www.jstor.org/stable/2950909>
- Sarafoglou, A., Kovacs, M., Bakos, B., Wagenmakers, E.-J., & Aczel, B. (2022). A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science*, 9 (7), 211997. <https://doi.org/10.1098/rsos.211997>
- Schaerer, M., du Plessis, C., Nguyen, M. H. B., van Aert, R. C. M., Tiokhin, L., Lakens, D., Clemente, E. G., Pfeiffer, T., Dreber, A., Johannesson, M., Clark, C. J., Gender Audits Forecasting Collaboration, & Uhlmann, E. L. (2023). On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions. *Organizational Behavior and Human Decision Processes*, 179, 104280. <https://doi.org/10.1016/j.obhdp.2023.104280>
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4), 551-566. <https://doi.org/10.1037/a0029487>
- Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Prasad, V. V., Bernstein, A., Robinson, E., Snellman, K., Sommer, S. A., Otner, S. M. G., Robinson, D., Madan, N., Silberzahn, R., Goldstein, P., Tierney, W., Murase, T., Mandl, B., Viganola, D., Strobl, C., Schaumans, C. B. C., Kelchtermans, S., Naseeb, C., Garrison, S. M., Yarkoni, T., Chan, C. S. R., Adie, P., Alaburda, P., Albers, C., Alspaugh, S., Alstott, J., Nelson, A. A., Ariño de la Rubia, E., Arzi, A., Bahník, Š., Baik, J., Balling, L. W., Banker, S., Aa Baranger, D., Barr, D. J., Barros-Rivera, B., Bauer, M., Blaise, E., Boelen, L., Carbonell, K. B., Briers, R. A., Burkhard, O., Canela, M.-A., Castrillo, L., Catlett, T.,

- Chen, O., Clark, M., Cohn, B., Coppock, A., Cugueró-Escofet, N., Curran, P. G., Cyrus-Lai, W., Dai, D., Dalla Riva, G. V., Danielsson, H., Russo, R. F. S. M., de Silva, N., Derungs, C., Dondelinger, F., Duarte de Souza, C., Dube, B. T., Dubova, M., Dunn, B. M., Edelsbrunner, P. A., Finley, S., Fox, N., Gnambs, T., Gong, Y., Grand, E., Greenawalt, B., Han, D., Hanel, P. H. P., Hong, A. B., Hood, D., Hsueh, J., Huang, L., Hui, K. N., Hultman, K. A., Javaid, A., Jiang, L. J., Jong, J., Kamdar, J., Kane, D., Kappler, G., Kaszubowski, E., Kavanagh, C. M., Khabisa, M., Kleinberg, B., Kouros, J., Krause, H., Kryptos, A.-M., Lavbič, D., Lee, R. L., Leffel, T., Lim, W. Y., Liverani, S., Loh, B., Lønsmann, D., Low, J. W., Lu, A., MacDonald, K., Madan, C. R., Madsen, L. H., Maimone, C., Mangold, A., Marshall, A., Matskewich, H. E., Mavon, K., McLain, K. L., McNamara, A. A., McNeill, M., Mertens, U., Miller, D., Moore, B., Moore, A., Nantz, E., Nasrullah, Z., Nejkovic, V., Nell, C. S., Nelson, A. A., Nilsonne, G., Nolan, R., O'Brien, C. E., O'Neill, P., O'Shea, K., Olita, T., Otterbacher, J., Palsetia, D., Pereira, B., Pozdniakov, I., Protzko, J., Reyt, J.-N., Riddle, T., Ridhwan Omar Ali, A., Ropovik, I., Rosenberg, J. M., Rothen, S., Schulte-Mecklenbeck, M., Sharma, N., Shotwell, G., Skarzynski, M., Stedden, W., Stodden, V., Stoffel, M. A., Stoltzman, S., Subbaiah, S., Tatman, R., Thibodeau, P. H., Tomkins, S., Valdivia, A., Druijff-van de Woestijne, G. B., Viana, L., Villesèche, F., Wadsworth, W. D., Wanders, F., Watts, K., Wells, J. D., Whelpley, C. E., Won, A., Wu, L., Yip, A., Youngflesh, C., Yu, J.-C., Zandian, A., Zhang, L., Zibman, C., & Uhlmann, E. L. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*, 165, 228-249. <https://doi.org/10.1016/j.obhdp.2021.02.003>
- Shrestha, Y. R., He, V. F., Puranam, P., & Krogh, G. (2021). Algorithm supported induction for Building Theory: How Can We Use Prediction Models to Theorize? *Organization Science*, 32(3), 856–880. <https://doi.org/10.1287/orsc.2020.1382>
- Silberzahn, R., Simonsohn, U., & Uhlmann, E. L. (2014). Matched-names analysis reveals no evidence of name-meaning effects: A collaborative commentary on Silberzahn and Uhlmann (2013). *Psychological Science*, 25(7), 1504-1505. <https://doi.org/10.1177/0956797614533802>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T. J., Hederos, K., Heene, M., Hofelich Mohr, A. J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D. M., Lei, R., Lindsay, T. A., Liverani, S., Madan, C. R., Molden, D., Molleman, E., Morey, R. D., Mulder, L. B., Nijstad, B. R., Pope, N. G., Pope, B., Prenoveau, J. M., Rink, F., Robusto, E., Roderique, H., Sandberg, A., Schlüter, E., Schönbrodt, F. D., Sherman, M. F., Sommer, S. A., Sotak, K., Spain, S., Spörlein, C., Stafford, T., Stefanutti, L., Tauber, S., Ullrich, J., Vianello, M., Wagenmakers, E.-J., Witkowiak, M., Yoon, S., & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337-356. <https://doi.org/10.1177/2515245917747646>
- Simonsohn, U. (2011a). Spurious also? Name-similarity effects (implicit egotism) in employment decisions. *Psychological Science*, 22(8), 1087-1089. <http://www.jstor.org/stable/25835504>

- Simonsohn, U. (2011b). Spurious? Name similarity effects (implicit egotism) in marriage, job, and moving decisions. *Journal of Personality and Social Psychology*, 101, 1-24.
<https://doi.org/10.1037/a0021990>
- Simonsohn, U. (2013). It really just does not follow, comments on Francis (2013). *Journal of Mathematical Psychology*, 57(5), 174-176. <https://doi.org/10.1016/j.jmp.2013.03.006>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143, 534-547. <https://doi.org/10.1037/a0033242>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208-1214. <https://doi.org/10.1038/s41562-020-0912-z>
- Smerdon, D., Hu, H., McLennan, A., von Hippel, W., & Albrecht, S. (2020). Female chess players show typical stereotype-threat effects: Commentary on Stafford (2018). *Psychological Science*, 31(6), 756-759. <https://doi.org/10.1177/0956797620924051>
- Stafford, T. (2018). Female chess players outperform expectations when playing men. *Psychological Science*, 29(3), 429-436. <https://doi.org/10.1177/0956797617736887>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712.
<https://doi.org/10.1177/1745691616658637>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632-638.
<https://doi.org/10.1177/1745691612463078>