Many Labs 2

Investigating Variation in Replicability across Sample and Setting

Richard Klein
LIP/PC2S
Université Grenoble Alpes

2018-12-12 (updated: 2018-12-12)

Many Labs 2

Replication Crisis

Replication Crisis

From cause for concern...

Replication Crisis

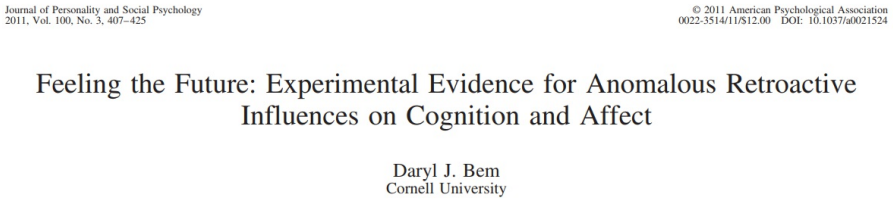
From cause for concern...

Open access, freely

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

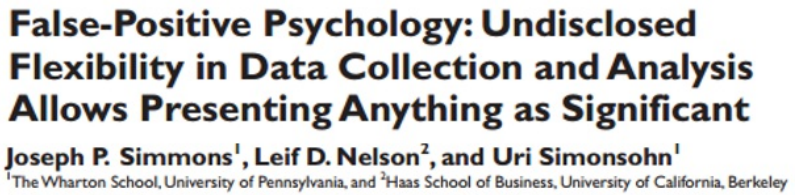


Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University



False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Replication Crisis

...to evidence of a problem...

- `Reproducibility Project: Psychology` (OSC, 2015)
  - ~40/100 studies replicated
- `Social Sciences Replication Project` (Camerer et al., 2018)
  - 13/21 replicated
- `Multiple large-scale Registered Reports`

Replication Crisis

...to addressing the problem

Replication Crisis

...to addressing the problem

- **What we know:**
  - It's easy to fool yourself with data ($p < .05$).

Replication Crisis

...to addressing the problem

- **What we know:**
  - It's easy to fool yourself with data ($p < .05$).
  - p values =/= truth.

Replication Crisis

...to addressing the problem

- **What we know:**
  - It's easy to fool yourself with data ($p < .05$).
  - p values =/= truth.
- **What we want to know:**
  - How to ensure our own results are replicable.

Replication Crisis

...to addressing the problem

- **What we know:**
  - It's easy to fool yourself with data ($p < .05$).
  - p values =/= truth.
- **What we want to know:**
  - How to ensure our own results are replicable.
- **What we don't know:**
  - Very much about replication.

Replication Crisis

...to addressing the problem

- **What we know:**
  - It's easy to fool yourself with data ($p < .05$).
  - p values =/= truth.
- **What we want to know:**
  - How to ensure our own results are replicable.
- **What we don't know:**
  - Very much about replication.
  - Must improve understanding to inform solutions

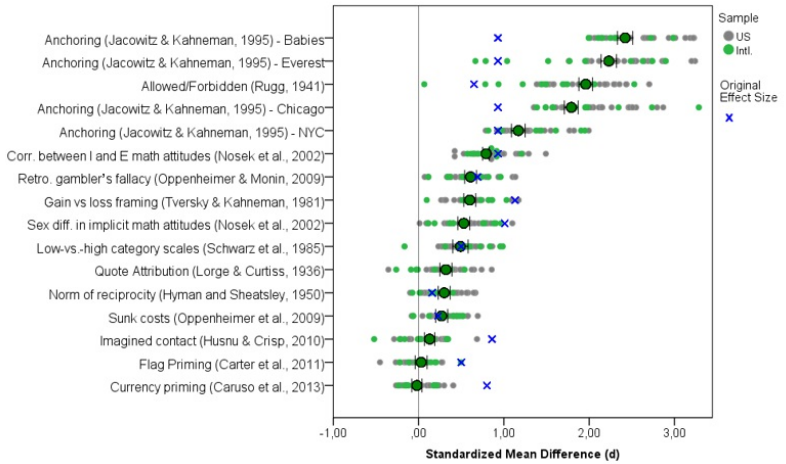
Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

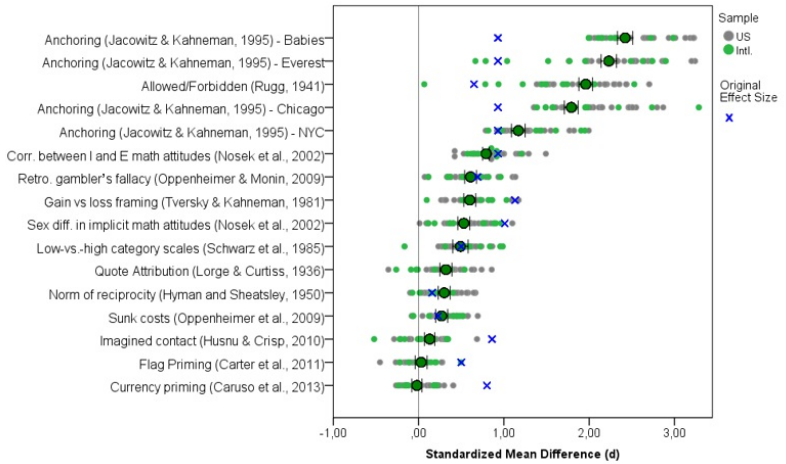
- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples



Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

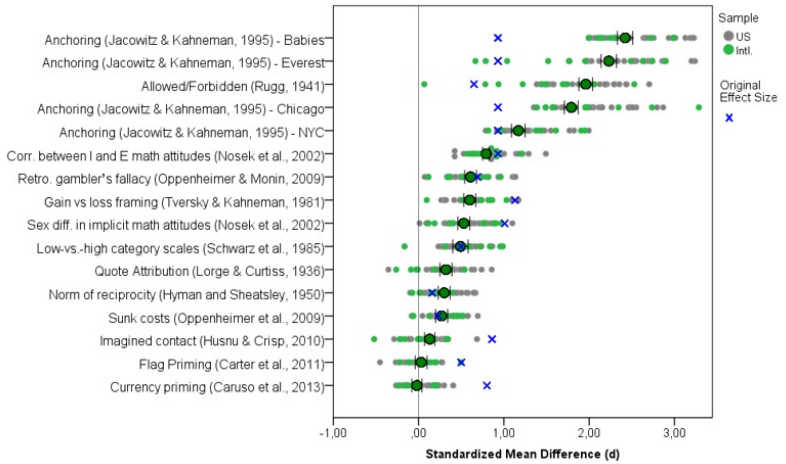
- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples
- `Many Labs 2` (Klein et al., in press)
  - Discussing today



Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

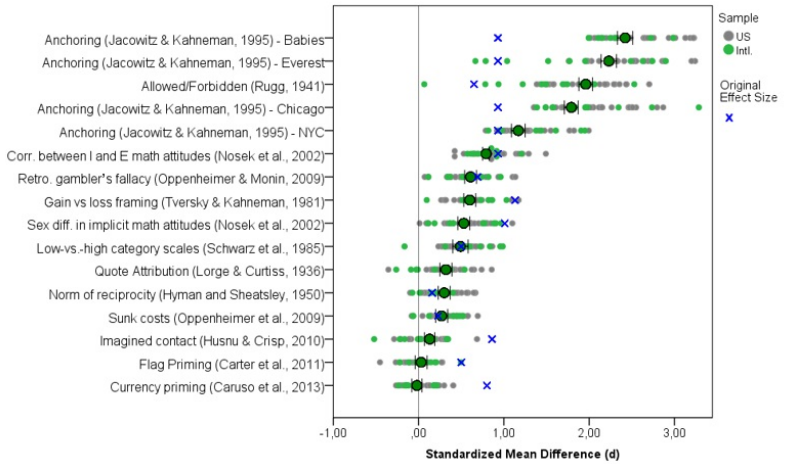
- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples
- `Many Labs 2` (Klein et al., in press)
  - Discussing today
- `Many Labs 3` (Ebersole et al., 2016)
  - 3/10 successful replications
  - Little variation across semester



Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

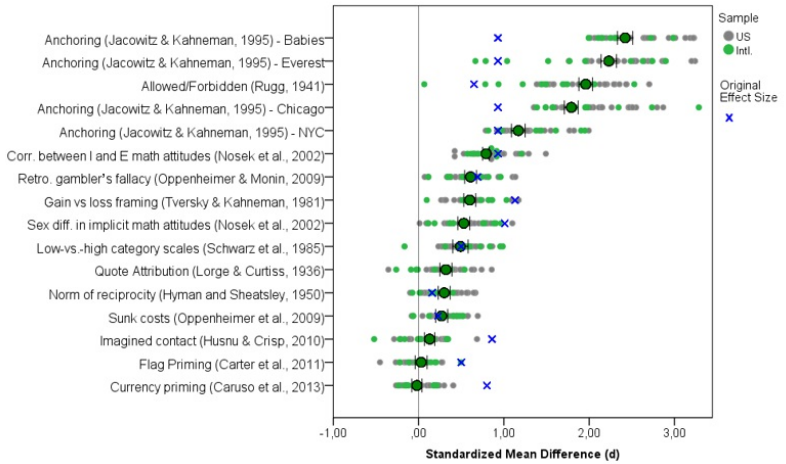
- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples
- `Many Labs 2` (Klein et al., in press)
  - Discussing today
- `Many Labs 3` (Ebersole et al., 2016)
  - 3/10 successful replications
  - Little variation across semester
- `Many Labs 4` (Klein et al., in prep)
  - Terror Management Theory-specific
  - Compare expert replications vs "in-house" replications



Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

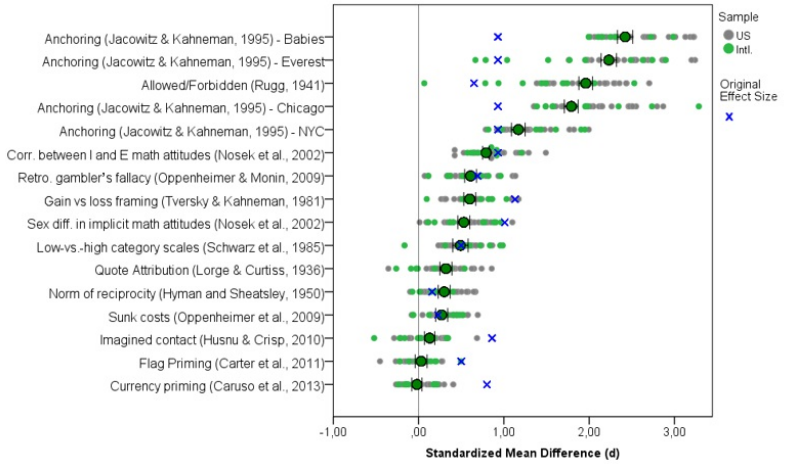
- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples
- `Many Labs 2` (Klein et al., in press)
  - Discussing today
- `Many Labs 3` (Ebersole et al., 2016)
  - 3/10 successful replications
  - Little variation across semester
- `Many Labs 4` (Klein et al., in prep)
  - Terror Management Theory-specific
  - Compare expert replications vs "in-house" replications
- `Many Labs 5` (Ebersole et al., in prep)
  - Follow-up to Reproducibility Project



Many Labs Projects

Large collaborations of researchers replicating the same findings. Each project examines a different aspect of replication.

- `Many Labs 1` (Klein et al., 2014)
  - 36 labs
  - 10/13 successful replications
  - Little variation between samples
- `Many Labs 2` (Klein et al., in press)
  - Discussing today
- `Many Labs 3` (Ebersole et al., 2016)
  - 3/10 successful replications
  - Little variation across semester
- `Many Labs 4` (Klein et al., in prep)
  - Terror Management Theory-specific
  - Compare expert replications vs "in-house" replications
- `Many Labs 5` (Ebersole et al., in prep)
  - Follow-up to Reproducibility Project
- Also: `Many Beds` (IJzerman et al.,), `Many Babies`, `Many Analysts`, `Many Labs VR`, `Many Numbers`, `Brazilian Reproducibility Project`



Many Labs 2

Many Labs 2

Like Many Labs 1, but a much stronger test:

Many Labs 2

Like Many Labs 1, but a much stronger test:

- **Question:** Will psychology studies replicate when they're repeated in contexts and samples very different from the original?

Many Labs 2

Like Many Labs 1, but a much stronger test:

- **Question:** Will psychology studies replicate when they're repeated in contexts and samples very different from the original?
- **Goal:** Replicate many different studies all around the world and compare if they vary based on the sample of data collection.

Many Labs 2

Like Many Labs 1, but a much stronger test:

- **Question:** Will psychology studies replicate when they're repeated in contexts and samples very different from the original?
- **Goal:** Replicate many different studies all around the world and compare if they vary based on the sample of data collection.
- Replicated 28 studies
  - Selected for impact, diversity of content, possibility for variation
  - Split across two study "packages" due to length
  - Computerized in Qualtrics
  - Randomized study order, presented back-to-back

Many Labs 2 Hsee example

Coats range from $100-$1000
Your friend buys you a $110 coat

Scarves range from $10-$100
Your friend buys you a $90 scarf

How generous was your friend?

Many Labs 2

Like Many Labs 1, but a much stronger test:

- **Question:** Will psychology studies replicate when they're repeated in contexts and samples very different from the original?
- **Goal:** Replicate many different studies all around the world and compare if they vary based on the sample of data collection.
- Replicated 28 studies
  - Selected for impact, diversity of content, possibility for variation
  - Split across two study "packages" due to length
  - Computerized in Qualtrics
  - Randomized study order, presented back-to-back

Many Labs 2

Like Many Labs 1, but a much stronger test:

- **Question:** Will psychology studies replicate when they're repeated in contexts and samples very different from the original?
- **Goal:** Replicate many different studies all around the world and compare if they vary based on the sample of data collection.
- Replicated 28 studies
  - Selected for impact, diversity of content, possibility for variation
  - Split across two study "packages" due to length
  - Computerized in Qualtrics
  - Randomized study order, presented back-to-back
- Administer identical study package across as many diverse samples as possible

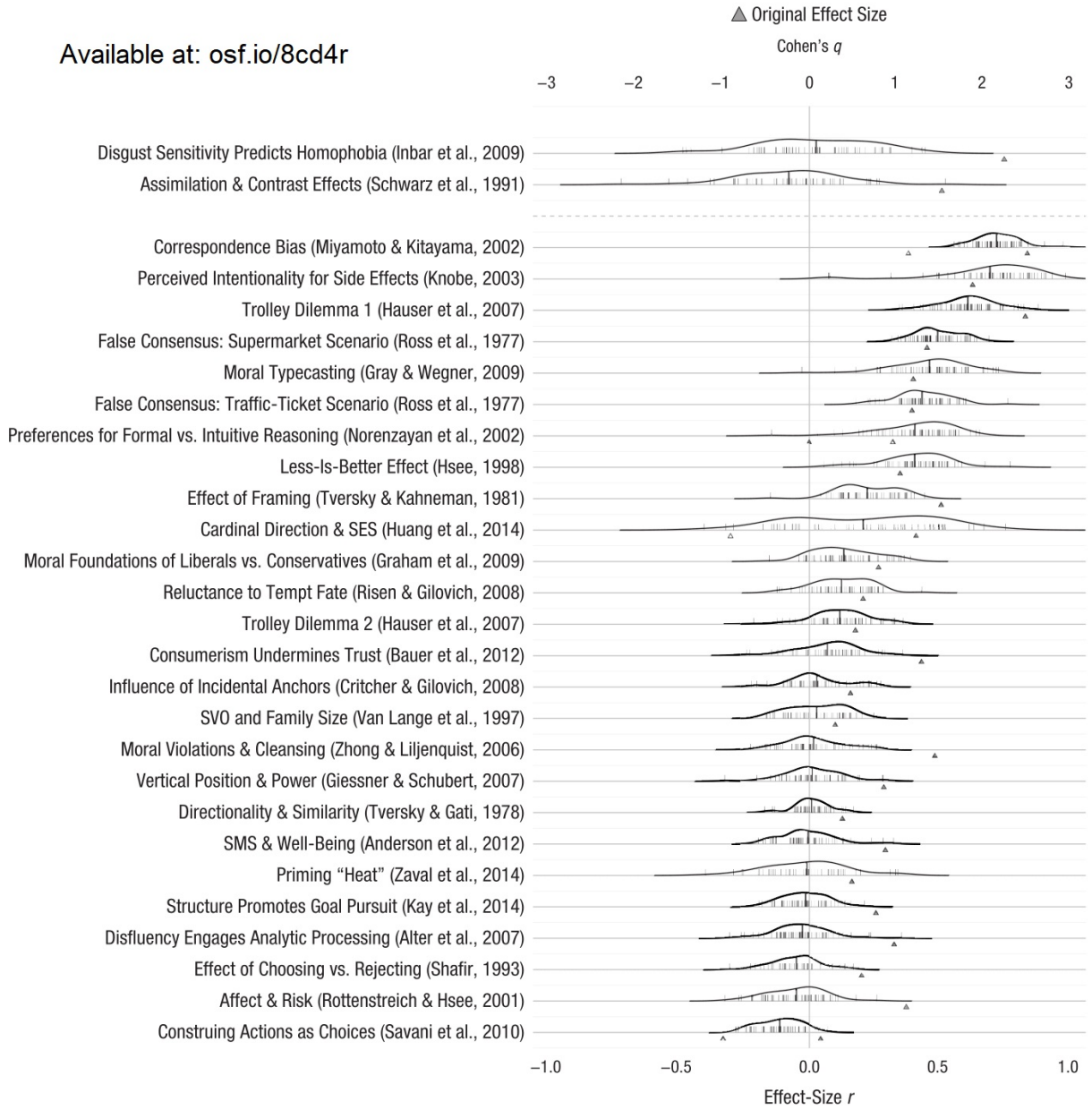
Many Labs 1 Map

Many Labs 2 Map

Many Labs 2

- 125 samples (each study administered in 60+)
- 36 countries, translated into 16 languages
- 15,305 participants total

Available at: osf.io/8cd4r



Many Labs 2 Results

Many Labs 2 Results

- 14/28 successful replications
  - $p < .0001$, non-trivial effect size, same direction as original
  - One additional weakly supported: $p = .03$

Many Labs 2 Results

- 14/28 successful replications
  - $p < .0001$, non-trivial effect size, same direction as original
  - One additional weakly supported: $p = .03$
- 21/28 had smaller effect size than original
  - Median original $d = 0.60$
  - Median replication $d = 0.15$

Many Labs 2 Heterogeneity

Many Labs 2 Heterogeneity

- Q statistic: 11/28 had $p < .001$ (statistically significant heterogeneity)
  - For 11 studies, observed variability across sites exceeded that which would be expected due to chance.

Many Labs 2 Heterogeneity

- Q statistic: 11/28 had $p < .001$ (statistically significant heterogeneity)
  - For 11 studies, observed variability across sites exceeded that which would be expected due to chance.
- However: 26/28 Tau ≤ 0.1
  - Variability across sites *existed*, but only had a very small effect (except for 1 or 2 studies)

Discussion

Discussion

- Low variation across sample/context

Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample

Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically

Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Replication rate aligns with other projects
  - Is this meaningful?

Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly (and robust replicability is a feasible goal)

Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly (and robust replicability is a feasible goal)
  - Reinforces need for solutions to ensure replicability

Discussion

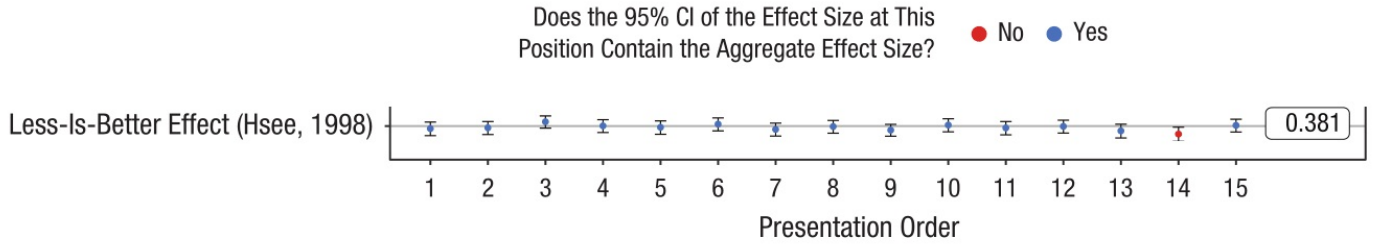
- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly (and robust replicability is a feasible goal)
  - Reinforces need for solutions to ensure replicability
- Open data: **https://osf.io/8cd4r/**
  - CC0, free use (any purpose)
  - We barely scratched surface

Thanks!

Special thanks to co-leads Fred Hasselman, Michelangelo Vianello, and Brian Nosek + 186 other co-authors.

Questions/comments?

@raklein3
raklein22@gmail.com
https://www.raklein.me


Does the 95% CI of the Effect Size at This Position Contain the Aggregate Effect Size?
No
Yes
Less-Is-Better Effect (Hsee, 1998)
0.381
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Presentation Order

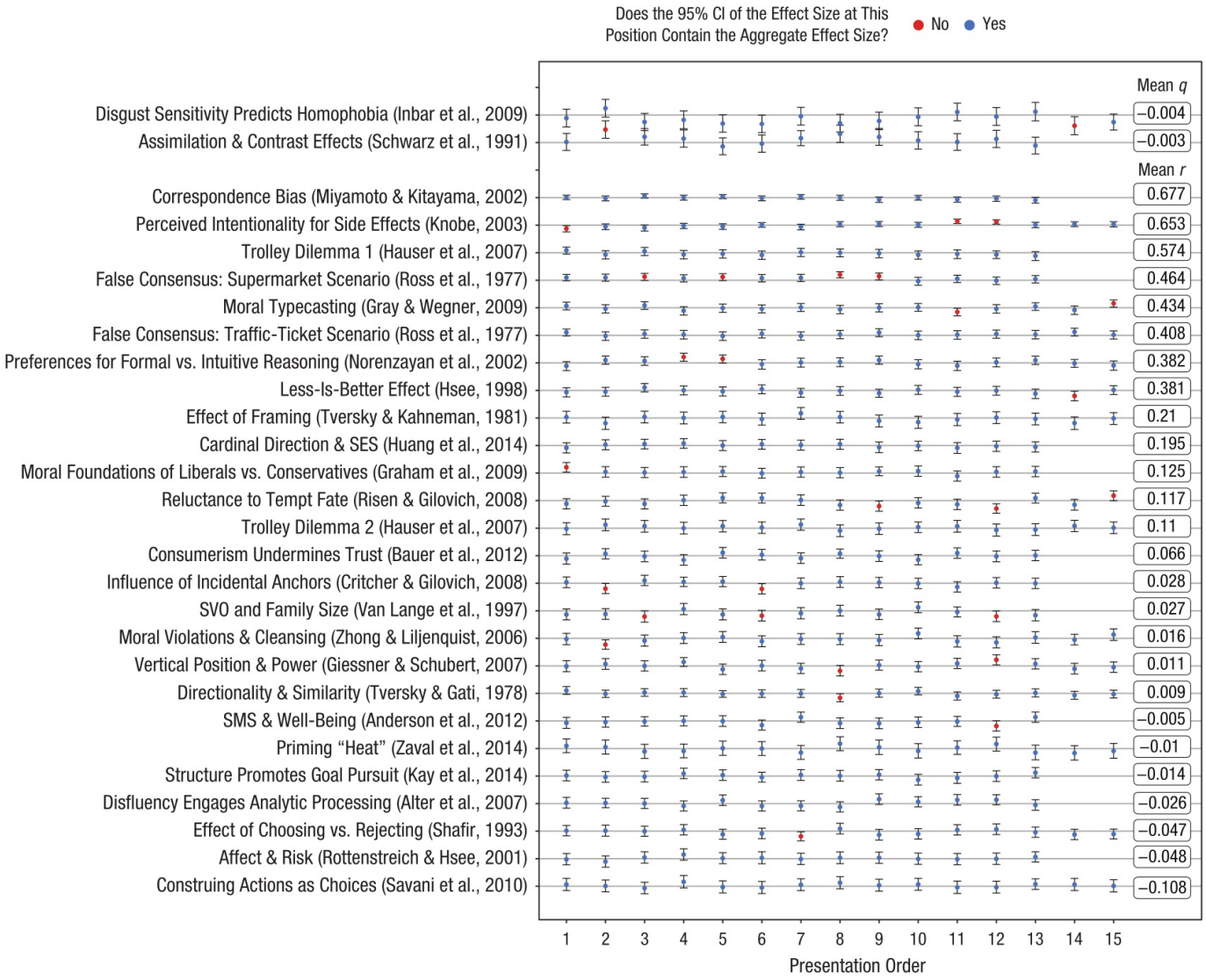

Does the 95% CI of the Effect Size at This Position Contain the Aggregate Effect Size?
No
Yes
Mean q
Disgust Sensitivity Predicts Homophobia (Inbar et al., 2009)
−0.004
Assimilation & Contrast Effects (Schwarz et al., 1991)
−0.003
Mean r
Correspondence Bias (Miyamoto & Kitayama, 2002)
0.677
Perceived Intentionality for Side Effects (Knobe, 2003)
0.653
Trolley Dilemma 1 (Hauser et al., 2007)
0.574
False Consensus: Supermarket Scenario (Ross et al., 1977)
0.464
Moral Typecasting (Gray & Wegner, 2009)
0.434
False Consensus: Traffic-Ticket Scenario (Ross et al., 1977)
0.408
Preferences for Formal vs. Intuitive Reasoning (Norenzayan et al., 2002)
0.382
Less-Is-Better Effect (Hsee, 1998)
0.381
Effect of Framing (Tversky & Kahneman, 1981)
0.21
Cardinal Direction & SES (Huang et al., 2014)
0.195
Moral Foundations of Liberals vs. Conservatives (Graham et al., 2009)
0.125
Reluctance to Tempt Fate (Risen & Gilovich, 2008)
0.117
Trolley Dilemma 2 (Hauser et al., 2007)
0.11
Consumerism Undermines Trust (Bauer et al., 2012)
0.066
Influence of Incidental Anchors (Critcher & Gilovich, 2008)
0.028
SVO and Family Size (Van Lange et al., 1997)
0.027
Moral Violations & Cleansing (Zhong & Liljenquist, 2006)
0.016
Vertical Position & Power (Giessner & Schubert, 2007)
0.011
Directionality & Similarity (Tversky & Gati, 1978)
0.009
SMS & Well-Being (Anderson et al., 2012)
−0.005
Priming “Heat” (Zaval et al., 2014)
−0.01
Structure Promotes Goal Pursuit (Kay et al., 2014)
−0.014
Disfluency Engages Analytic Processing (Alter et al., 2007)
−0.026
Effect of Choosing vs. Rejecting (Shafir, 1993)
−0.047
Affect & Risk (Rottenstreich & Hsee, 2001)
−0.048
Construing Actions as Choices (Savani et al., 2010)
−0.108
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Presentation Order

Table 3. Results of Heterogeneity Tests for Each of the 28 Effects

Effect	ES[a]	All samples (no moderators) Tau	*Q*	*df*	*p*	I^2
						Cohen's *q* effect size
Disgust sensitivity predicts homophobia (Inbar, Pizarro, Knobe, & Bloom, 2009)	0.05	.00	55.80	58.00	.56	3.00% [0%, 30%]
Assimilation and contrast effects in question sequences (Schwarz, Strack, & Mai, 1991)	−0.07	.10	60.39	58.00	.39	15.00% [0%, 33%]
						Cohen's *d* effect size
Correspondence bias (Miyamoto & Kitayama, 2002)	1.82	.00	235.65	57.00	< .001	65.00% [46%, 73%]
Perceived intentionality for side effects (Knobe, 2003)	1.75	.14	631.72	58.00	< .001	93.00% [92%, 97%]
Trolley Dilemma 1: principle of double effect (Hauser, Cushman, Young, Jin, & Mikhail, 2007)	1.35	.10	131.24	58.00	< .001	54.00% [32%, 66%]
False Consensus: supermarket scenario (Ross, Greene, & House, 1977)	1.18	.00	65.54	58.00	.23	16.00% [0%, 41%]
Moral typecasting (Gray & Wegner, 2009)	0.95	.10	203.30	59.00	< .001	73.00% [62%, 83%]
False Consensus: traffic-ticket scenario (Ross et al., 1977)	0.95	.00	100.19	57.00	< .001	43.00% [18%, 62%]
Preferences for formal versus intuitive reasoning (Norenzayan, Smith, Kim, & Nisbett, 2002)	0.86	.10	156.75	56.00	< .001	66.00% [54%, 81%]
Less is better (Hsee, 1998)	0.78	.10	158.41	56.00	< .001	65.00% [49%, 77%]
Effect of framing on decision making (Tversky & Kahneman, 1981)	0.40	.00	55.20	54.00	.43	6.00% [0%, 36%]
Cardinal direction and socioeconomic status (Huang, Tse, & Cho, 2014)	0.40	.24	626.26	63.00	< .001	89.00% [84%, 92%]
Moral foundations of liberals versus conservatives (Graham, Haidt, & Nosek, 2009)	0.29	.09	175.26	59.00	< .001	64.00% [49%, 75%]
Reluctance to tempt fate (Risen & Gilovich, 2008)	0.18	.00	87.82	58.00	.01	36.00% [6%, 54%]

Table 3. (Continued)

Effect	ES[a]	All samples (no moderators) Tau	*Q*	*df*	*p*	I^2
Trolley Dilemma 2: principle of double effect (Hauser et al., 2007)	0.25	.00	60.40	59.00	.42	12.00% [0%, 33%]
Consumerism undermines trust (Bauer, Wilkie, Kim, & Bodenhausen, 2012)	0.12	.00	63.78	53.00	.15	12.00% [0%, 49%]
Influence of incidental anchors on judgment (Critcher & Gilovich, 2008)	0.04	.00	64.88	58.00	.25	6.00% [0%, 43%]
Social value orientation and family size (Van Lange, Otten, De Bruin, & Joireman, 1997)	−0.03	.00	103.56	53.00	< .001	50.00% [28%, 68%]
Moral violations and desire for cleansing (Zhong & Liljenquist, 2006)	0.00	.00	65.59	51.00	.08	22.00% [0%, 52%]
Vertical position and power (Giessner & Schubert, 2007)	0.03	.00	62.87	58.00	.31	*3.00% [0%, 42%]*
Directionality and similarity (Tversky & Gati, 1978)	0.01	.00	15.33	48.00	.99	0.00% [0%, 0%]
Sociometric status and well-being (Anderson, Kraus, Galinsky, & Keltner, 2012)	−0.04	.00	55.09	58.00	.58	2.00% [0%, 30%]
Priming "heat" increases belief in global warming (Zaval, Keenan, Johnson, & Weber, 2014)	−0.03	.10	72.96	46.00	.01	37.00% [8%, 63%]
Structure promotes goal pursuit (Kay, Laurin, Fitzsimons, & Landau, 2014)	−0.02	.00	33.95	51.00	.97	0.00% [0%, 2%]
Disfluency engages analytic processing (Alter, Oppenheimer, Epley, & Eyre, 2007)	−0.03	.00	59.46	65.00	.67	0.00% [0%, 27%]
Effect of choosing versus rejecting on relative desirability (Shafir, 1993)	−0.13	.00	51.67	40.00	.10	26.00% [0%, 52%]
Affect and risk (Rottenstreich & Hsee, 2001)	−0.08	.00	50.75	59.00	.77	0.00% [0%, 21%]
Construing actions as choices (Savani, Markus, Naidu, Kumar, & Berlia, 2010)	−0.18	.00	155.49	56.00	< .001	64.00% [47%, 76%]

Table 3. Results of Heterogeneity Tests for Each of the 28 Effects

Effect	ES[a]	All samples (no moderators) Tau	*Q*	*df*	*p*	I^2
						Cohen's *q* effect size
Disgust sensitivity predicts homophobia (Inbar, Pizarro, Knobe, & Bloom, 2009)	0.05	.00	55.80	58.00	.56	3.00% [0%, 30%]
Assimilation and contrast effects in question sequences (Schwarz, Strack, & Mai, 1991)	−0.07	.10	60.39	58.00	.39	15.00% [0%, 33%]
						Cohen's *d* effect size
Correspondence bias (Miyamoto & Kitayama, 2002)	1.82	.00	235.65	57.00	< .001	65.00% [46%, 73%]
Perceived intentionality for side effects (Knobe, 2003)	1.75	.14	631.72	58.00	< .001	93.00% [92%, 97%]
Trolley Dilemma 1: principle of double effect (Hauser, Cushman, Young, Jin, & Mikhail, 2007)	1.35	.10	131.24	58.00	< .001	54.00% [32%, 66%]
False Consensus: supermarket scenario (Ross, Greene, & House, 1977)	1.18	.00	65.54	58.00	.23	16.00% [0%, 41%]
Moral typecasting (Gray & Wegner, 2009)	0.95	.10	203.30	59.00	< .001	73.00% [62%, 83%]
False Consensus: traffic-ticket scenario (Ross et al., 1977)	0.95	.00	100.19	57.00	< .001	43.00% [18%, 62%]
Preferences for formal versus intuitive reasoning (Norenzayan, Smith, Kim, & Nisbett, 2002)	0.86	.10	156.75	56.00	< .001	66.00% [54%, 81%]
Less is better (Hsee, 1998)	0.78	.10	158.41	56.00	< .001	65.00% [49%, 77%]
Effect of framing on decision making (Tversky & Kahneman, 1981)	0.40	.00	55.20	54.00	.43	6.00% [0%, 36%]
Cardinal direction and socioeconomic status (Huang, Tse, & Cho, 2014)	0.40	.24	626.26	63.00	< .001	89.00% [84%, 92%]
Moral foundations of liberals versus conservatives (Graham, Haidt, & Nosek, 2009)	0.29	.09	175.26	59.00	< .001	64.00% [49%, 75%]
Reluctance to tempt fate (Risen & Gilovich, 2008)	0.18	.00	87.82	58.00	.01	36.00% [6%, 54%]

Table 3. (Continued)

Effect	ES[a]	All samples (no moderators) Tau	*Q*	*df*	*p*	I^2
Trolley Dilemma 2: principle of double effect (Hauser et al., 2007)	0.25	.00	60.40	59.00	.42	12.00% [0%, 33%]
Consumerism undermines trust (Bauer, Wilkie, Kim, & Bodenhausen, 2012)	0.12	.00	63.78	53.00	.15	12.00% [0%, 49%]
Influence of incidental anchors on judgment (Critcher & Gilovich, 2008)	0.04	.00	64.88	58.00	.25	6.00% [0%, 43%]
Social value orientation and family size (Van Lange, Otten, De Bruin, & Joireman, 1997)	−0.03	.00	103.56	53.00	< .001	50.00% [28%, 68%]
Moral violations and desire for cleansing (Zhong & Liljenquist, 2006)	0.00	.00	65.59	51.00	.08	22.00% [0%, 52%]
Vertical position and power (Giessner & Schubert, 2007)	0.03	.00	62.87	58.00	.31	*3.00% [0%, 42%]*
Directionality and similarity (Tversky & Gati, 1978)	0.01	.00	15.33	48.00	.99	0.00% [0%, 0%]
Sociometric status and well-being (Anderson, Kraus, Galinsky, & Keltner, 2012)	−0.04	.00	55.09	58.00	.58	2.00% [0%, 30%]
Priming "heat" increases belief in global warming (Zaval, Keenan, Johnson, & Weber, 2014)	−0.03	.10	72.96	46.00	.01	37.00% [8%, 63%]
Structure promotes goal pursuit (Kay, Laurin, Fitzsimons, & Landau, 2014)	−0.02	.00	33.95	51.00	.97	0.00% [0%, 2%]
Disfluency engages analytic processing (Alter, Oppenheimer, Epley, & Eyre, 2007)	−0.03	.00	59.46	65.00	.67	0.00% [0%, 27%]
Effect of choosing versus rejecting on relative desirability (Shafir, 1993)	−0.13	.00	51.67	40.00	.10	26.00% [0%, 52%]
Affect and risk (Rottenstreich & Hsee, 2001)	−0.08	.00	50.75	59.00	.77	0.00% [0%, 21%]
Construing actions as choices (Savani, Markus, Naidu, Kumar, & Berlia, 2010)	−0.18	.00	155.49	56.00	< .001	64.00% [47%, 76%]

Table 3. Results of Heterogeneity Tests for Each of the 28 Effects

Effect	ES[a]	All samples (no moderators) Tau	Q	df	p	I^2
						Cohen's *q* effect size
Disgust sensitivity predicts homophobia (Inbar, Pizarro, Knobe, & Bloom, 2009)	0.05	.00	55.80	58.00	.56	3.00% [0%, 30%]
Assimilation and contrast effects in question sequences (Schwarz, Strack, & Mai, 1991)	−0.07	.10	60.39	58.00	.39	15.00% [0%, 33%]
						Cohen's *d* effect size
Correspondence bias (Miyamoto & Kitayama, 2002)	1.82	.00	235.65	57.00	< .001	65.00% [46%, 73%]
Perceived intentionality for side effects (Knobe, 2003)	1.75	.14	631.72	58.00	< .001	93.00% [92%, 97%]
Trolley Dilemma 1: principle of double effect (Hauser, Cushman, Young, Jin, & Mikhail, 2007)	1.35	.10	131.24	58.00	< .001	54.00% [32%, 66%]
False Consensus: supermarket scenario (Ross, Greene, & House, 1977)	1.18	.00	65.54	58.00	.23	16.00% [0%, 41%]
Moral typecasting (Gray & Wegner, 2009)	0.95	.10	203.30	59.00	< .001	73.00% [62%, 83%]
False Consensus: traffic-ticket scenario (Ross et al., 1977)	0.95	.00	100.19	57.00	< .001	43.00% [18%, 62%]
Preferences for formal versus intuitive reasoning (Norenzayan, Smith, Kim, & Nisbett, 2002)	0.86	.10	156.75	56.00	< .001	66.00% [54%, 81%]
Less is better (Hsee, 1998)	0.78	.10	158.41	56.00	< .001	65.00% [49%, 77%]
Effect of framing on decision making (Tversky & Kahneman, 1981)	0.40	.00	55.20	54.00	.43	6.00% [0%, 36%]
Cardinal direction and socioeconomic status (Huang, Tse, & Cho, 2014)	0.40	.24	626.26	63.00	< .001	89.00% [84%, 92%]
Moral foundations of liberals versus conservatives (Graham, Haidt, & Nosek, 2009)	0.29	.09	175.26	59.00	< .001	64.00% [49%, 75%]
Reluctance to tempt fate (Risen & Gilovich, 2008)	0.18	.00	87.82	58.00	.01	36.00% [6%, 54%]

Table 3. (Continued)

Effect	ES[a]	All samples (no moderators) Tau	Q	df	p	I^2
Trolley Dilemma 2: principle of double effect (Hauser et al., 2007)	0.25	.00	60.40	59.00	.42	12.00% [0%, 33%]
Consumerism undermines trust (Bauer, Wilkie, Kim, & Bodenhausen, 2012)	0.12	.00	63.78	53.00	.15	12.00% [0%, 49%]
Influence of incidental anchors on judgment (Critcher & Gilovich, 2008)	0.04	.00	64.88	58.00	.25	6.00% [0%, 43%]
Social value orientation and family size (Van Lange, Otten, De Bruin, & Joireman, 1997)	−0.03	.00	103.56	53.00	< .001	50.00% [28%, 68%]
Moral violations and desire for cleansing (Zhong & Liljenquist, 2006)	0.00	.00	65.59	51.00	.08	22.00% [0%, 52%]
Vertical position and power (Giessner & Schubert, 2007)	0.03	.00	62.87	58.00	.31	*3.00% [0%, 42%]*
Directionality and similarity (Tversky & Gati, 1978)	0.01	.00	15.33	48.00	.99	0.00% [0%, 0%]
Sociometric status and well-being (Anderson, Kraus, Galinsky, & Keltner, 2012)	−0.04	.00	55.09	58.00	.58	2.00% [0%, 30%]
Priming "heat" increases belief in global warming (Zaval, Keenan, Johnson, & Weber, 2014)	−0.03	.10	72.96	46.00	.01	37.00% [8%, 63%]
Structure promotes goal pursuit (Kay, Laurin, Fitzsimons, & Landau, 2014)	−0.02	.00	33.95	51.00	.97	0.00% [0%, 2%]
Disfluency engages analytic processing (Alter, Oppenheimer, Epley, & Eyre, 2007)	−0.03	.00	59.46	65.00	.67	0.00% [0%, 27%]
Effect of choosing versus rejecting on relative desirability (Shafir, 1993)	−0.13	.00	51.67	40.00	.10	26.00% [0%, 52%]
Affect and risk (Rottenstreich & Hsee, 2001)	−0.08	.00	50.75	59.00	.77	0.00% [0%, 21%]
Construing actions as choices (Savani, Markus, Naidu, Kumar, & Berlia, 2010)	−0.18	.00	155.49	56.00	< .001	64.00% [47%, 76%]