

# Many Labs 2

## Investigating Variation in Replicability across Sample and Setting

Richard Klein

LIP/PC2S

Université Grenoble Alpes

2018-12-10 (updated: 2018-12-10)

# Many Labs 2

# Cause for Concern

# Cause for Concern

*Open access, freely*

Essay

## Why Most Published Research Findings Are False

John P. A. Ioannidis

# 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

© 2011 American Psychological Association  
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

# 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

© 2011 American Psychological Association  
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

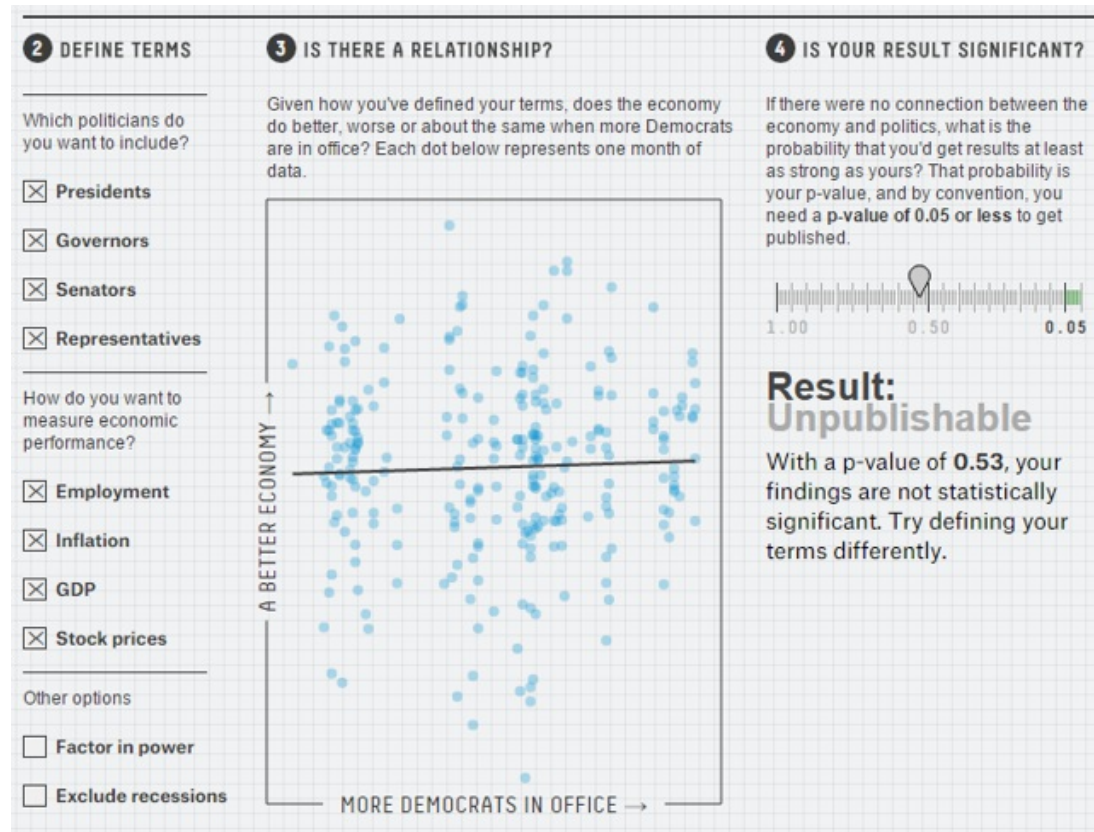
## False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons<sup>1</sup>, Leif D. Nelson<sup>2</sup>, and Uri Simonsohn<sup>1</sup>

<sup>1</sup>The Wharton School, University of Pennsylvania, and <sup>2</sup>Haas School of Business, University of California, Berkeley

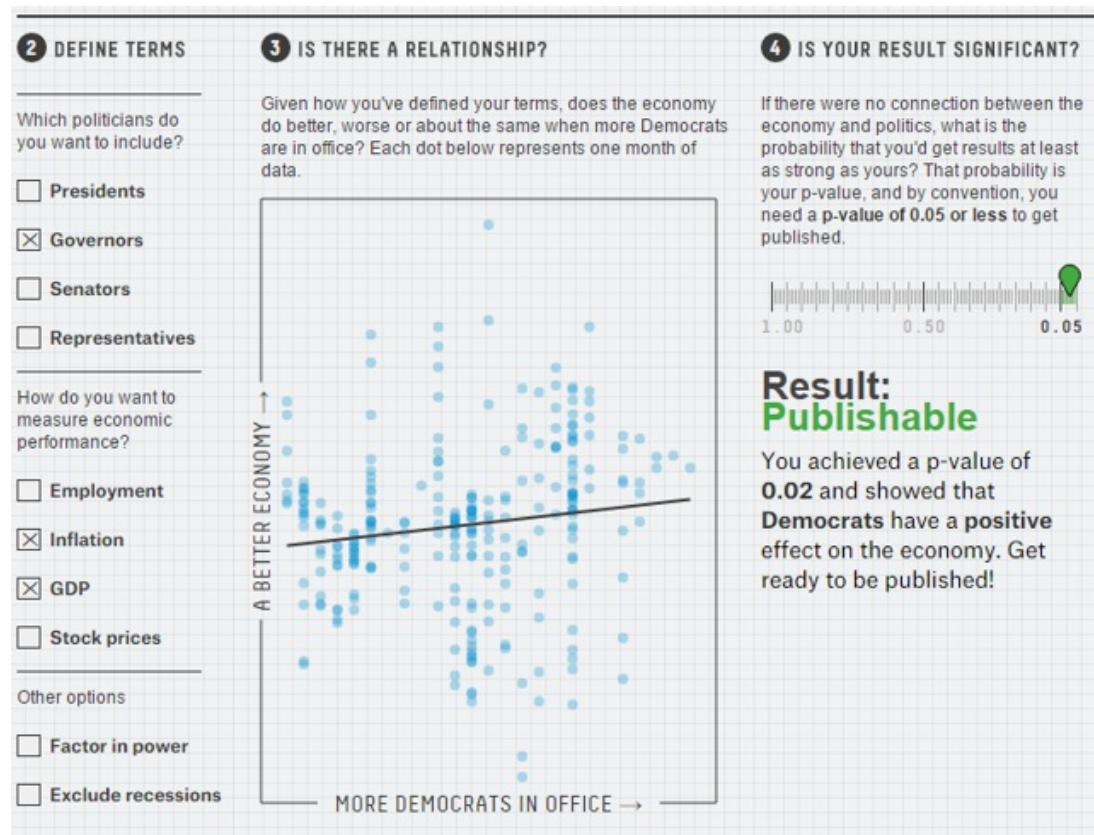
# Flexibility in Data Analysis

<http://fivethirtyeight.com/features/science-isnt-broken>



# Flexibility in Data Analysis

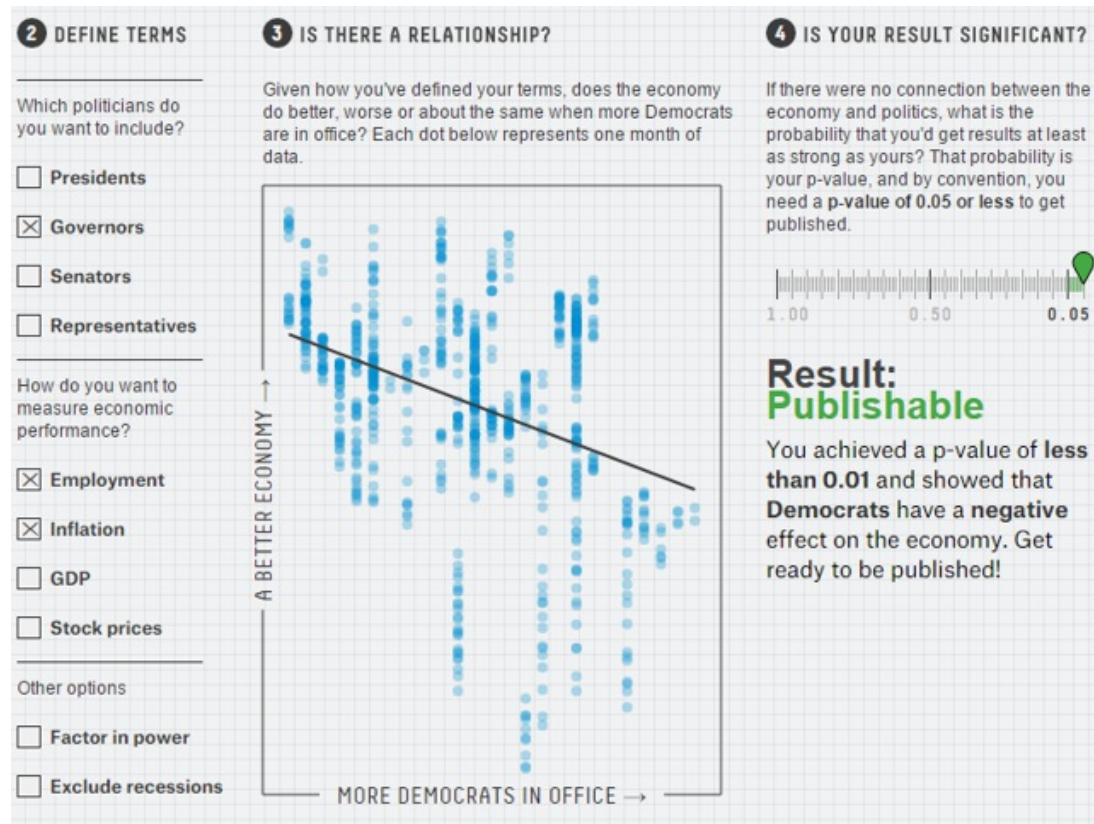
<http://fivethirtyeight.com/features/science-isnt-broken>





# Flexibility in Data Analysis

<http://fivethirtyeight.com/features/science-isnt-broken>



# Failures to Replicate

# Failures to Replicate

- **Reproducibility Project: Psychology** (OSC, 2015)
  - ~40/100 studies replicated

# Failures to Replicate

- **Reproducibility Project: Psychology** (OSC, 2015)
  - ~40/100 studies replicated
- **Social Sciences Replication Project** (Camerer et al., 2018)
  - 13/21 replicated
  - All from Science and Nature

# Failures to Replicate

- **Reproducibility Project: Psychology** (OSC, 2015)
  - ~40/100 studies replicated
- **Social Sciences Replication Project** (Camerer et al., 2018)
  - 13/21 replicated
  - All from Science and Nature
- **Multiple large-scale Registered Reports**
  - POPS/AMPPS Registered Replication Reports

# Failures to Replicate

# Failures to Replicate

- What we know: Many studies are failing to replicate

# Failures to Replicate

- What we know: Many studies are failing to replicate
- Why? Not sure



# Failures to Replicate

- What we know: Many studies are failing to replicate
- Why? Not sure
  - Could be false positives

# Failures to Replicate

- What we know: Many studies are failing to replicate
- Why? Not sure
  - Could be false positives
  - Could be many other reasons:
  - Moderators (known/unknown)
  - Lack of care/expertise
  - Sensitivity of effects to sample/context

# Many Labs Projects

# Many Labs Projects

Each ML project examines a different aspect of replication. Each question requires data collection at multiple labs.

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

# Many Labs Projects

Each ML project examines a different aspect of replication. Each question requires data collection at multiple labs.

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

# Many Labs Projects

Each ML project examines a different aspect of replication. Each question requires data collection at multiple labs.

- **Many Labs 1** (Klein et al., 2014)
  - 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

Each ML project examines a different aspect of replication. Each question requires data collection at multiple labs.

- **Many Labs 1** (Klein et al., 2014)
  - 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

Each ML project examines a different aspect of replication. Each question requires data collection at multiple labs.

- **Many Labs 1** (Klein et al., 2014)
  - 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 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:

- 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:

- Goal: Replicate many different studies all around the world and compare if they vary based on the sample of data collection.
- Replicated 28 studies
  - 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:

- Goal: Replicate many different studies all around the world and compare if they vary based on the sample of data collection.
- Replicated 28 studies
  - Split across two study "packages" due to length
  - Computerized in Qualtrics
  - Randomized study order, presented back-to-back
- Which studies?
  - Structured selection process by committee. Documented: [osf.io/8cd4r](https://osf.io/8cd4r)
  - Sought open nominations for studies
  - Emphasized impact (citations, etc.), diversity of content, possibility for variability across sites
  - But substantial practical constraints: Short, able to be computerized
  - Authors could decline to be replicated

# Many Labs 2

# Many Labs 2

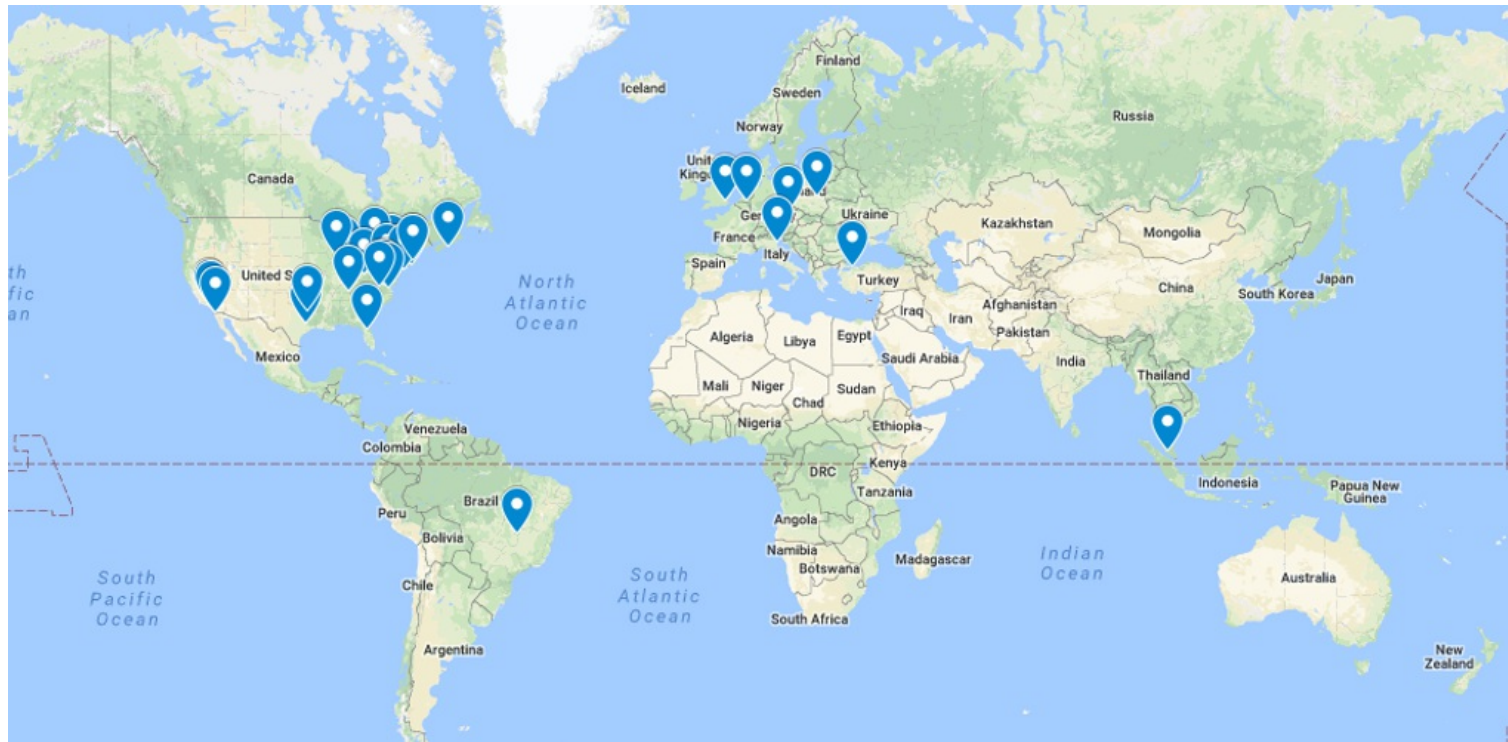
- Registered Replication Report at AMPPS:
  - Each study reviewed and approved by original authors or other experts
  - Analysis plan(s) specified in advance ([osf.io/c97pd/](https://osf.io/c97pd/))
  - Open data and materials

# Many Labs 2

- Registered Replication Report at AMPPS:
  - Each study reviewed and approved by original authors or other experts
  - Analysis plan(s) specified in advance ([osf.io/c97pd/](https://osf.io/c97pd/))
  - Open data and materials
- Administer packages across 125 samples
  - Slate 1: 13 studies administered in each of 61 labs
  - Slate 2: 15 studies administered in each of 64 labs
  - Sites (mostly) randomly assigned to slates
  - Minimum of 80 participants per site
  - 15,305 participants total
  - Much more diverse



# Many Labs 1 Map



# Many Labs 2 Map



# Many Labs 2 Hsee example



Coats range from \$100-\$1000

Your friend buys you a \$110 coat

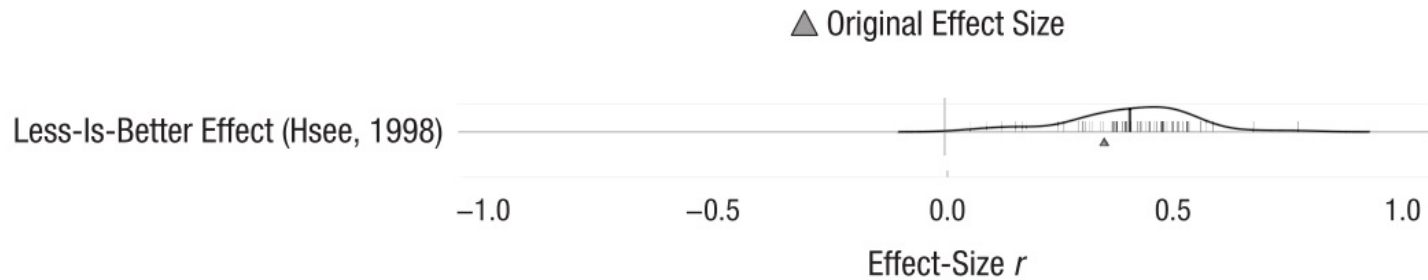


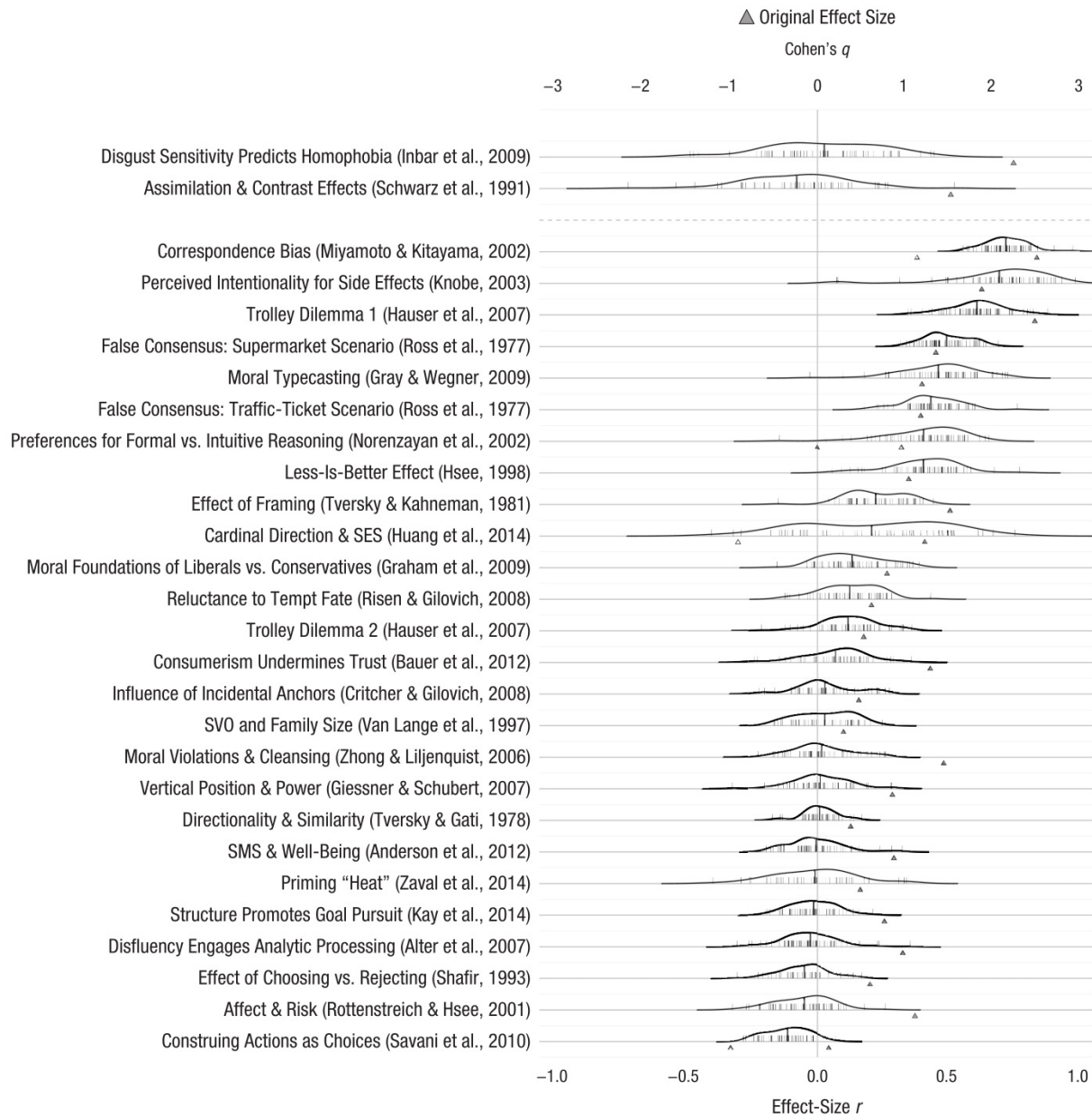
Scarves range from \$10-\$100

Your friend buys you a \$90 scarf

How generous was your friend?

# Many Labs 2 Hsee results





# Many Labs 2 Results

# Many Labs 2 Results

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

# Many Labs 2 Results

- 14/28 successful replications
  - $p < .0001$ , non-trivial effect size, same direction as original
  - One weakly supported,  $p = .03$  but near-zero effect size
- 75% has smaller effect size than original
  - Median original  $d = 0.60$
  - Median replication  $d = 0.15$

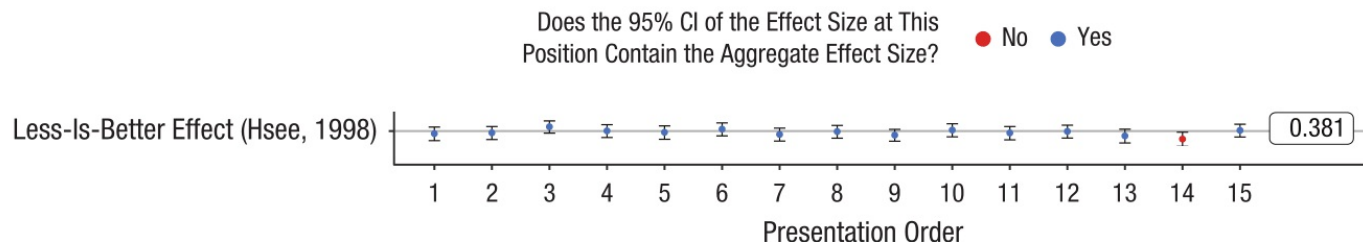


# Many Labs 2 Results

- 14/28 successful replications
  - $p < .0001$ , non-trivial effect size, same direction as original
  - One weakly supported,  $p = .03$  but near-zero effect size
- 75% has smaller effect size than original
  - Median original  $d = 0.60$
  - Median replication  $d = 0.15$
- No evidence that the order of the studies mattered
  - In general, didn't matter if the study came first, last, or in any other position.
  - Same as ML1 and ML3

# Many Labs 2 Results

- 14/28 successful replications
  - $p < .0001$ , non-trivial effect size, same direction as original
  - One weakly supported,  $p = .03$  but near-zero effect size
- 75% has smaller effect size than original
  - Median original  $d = 0.60$
  - Median replication  $d = 0.15$
- No evidence that the order of the studies mattered
  - In general, didn't matter if the study came first, last, or in any other position.
  - Same as ML1 and ML3





# Many Labs 2 Heterogeneity

# Many Labs 2 Heterogeneity

- Q statistic: (~ significance test for variation across sites exceeding chance)
  - 11/28 (39%) showed significant heterogeneity
  - Nearly all from larger-effects studies

# Many Labs 2 Heterogeneity

- Q statistic: (~ significance test for variation across sites exceeding chance)
  - 11/28 (39%) showed significant heterogeneity
  - Nearly all from larger-effects studies
- $I^2$ :
  - 36% showed at least medium heterogeneity
  - Likely not an appropriate measure in this case:
  - See [osf.io/frbuw](https://osf.io/frbuw) (Marcel van Assen), [Datacolada.org/63](https://datacolada.org/63), Borenstein+Higgins

# Many Labs 2 Heterogeneity

- Q statistic: (~ significance test for variation across sites exceeding chance)
  - 11/28 (39%) showed significant heterogeneity
  - Nearly all from larger-effects studies
- $I^2$ :
  - 36% showed at least medium heterogeneity
  - Likely not an appropriate measure in this case:
  - See [osf.io/frbuw](https://osf.io/frbuw) (Marcel van Assen), [Datacolada.org/63](https://datacolada.org/63), Borenstein+Higgins
- Tau is probably best
  - SD across samples in the unit of the effect size (after accounting for sampling error)

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

Effect	ES <sup>a</sup>	All samples (no moderators)				
		Tau	<i>Q</i>	<i>df</i>	<i>p</i>	<i>I</i> <sup>2</sup>
						Cohen's <i>q</i> 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 <i>d</i> 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 <sup>a</sup>	All samples (no moderators)				
		Tau	<i>Q</i>	<i>df</i>	<i>p</i>	<i>I</i> <sup>2</sup>
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, Naik, 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 <sup>a</sup>	Tau	All samples (no moderators)			
			<i>Q</i>	<i>df</i>	<i>p</i>	<i>I</i> <sup>2</sup>
						Cohen's <i>q</i> 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 <i>d</i> 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 <sup>a</sup>	Tau	All samples (no moderators)			
			<i>Q</i>	<i>df</i>	<i>p</i>	<i>I</i> <sup>2</sup>
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, Naik, 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 <sup>a</sup>	All samples (no moderators)				
		Tau	Q	df	p	I <sup>2</sup>
						Cohen's <i>q</i> 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 <i>d</i> 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 <sup>a</sup>	All samples (no moderators)				
		Tau	Q	df	p	I <sup>2</sup>
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, Naik, Kumar, & Berlia, 2010)	-0.18	.00	155.49	56.00	< .001	64.00% [47%, 76%]

# 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
- Big caveat: Mostly student samples, mostly short computerized studies

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- 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
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly
- Personal takeaways:

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly
- Personal takeaways:
  - Robust replicability is a feasible goal (for many studies)

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly
- Personal takeaways:
  - Robust replicability is a feasible goal (for many studies)
  - Nudges me towards "false-positive" explanation for replication failures (in general)

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly
- Personal takeaways:
  - Robust replicability is a feasible goal (for many studies)
  - Nudges me towards "false-positive" explanation for replication failures (in general)
  - Reinforces need for preregistration/Registered Reports

# Discussion

- Low variation across sample/context
  - Not reasonable to discount replications *by default* based on sample
  - Instead, test moderators empirically
- Big caveat: Mostly student samples, mostly short computerized studies
- Replication rate aligns with other projects
  - Is this meaningful?
- Many studies replicate robustly
- Personal takeaways:
  - Robust replicability is a feasible goal (for many studies)
  - Nudges me towards "false-positive" explanation for replication failures (in general)
  - Reinforces need for preregistration/Registered Reports
- 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>



financed by  
**IDEX Université Grenoble Alpes**