

Open Science Toolbox:

Leveraging Methodological Reforms to Accelerate Research

ICPS Paris March 2019

ALL SLIDES: <https://osf.io/tve2q/>

2019-03-08 (updated: 2019-03-09)

Many Labs 2

Investigating Variation in Replicability across Sample and
Setting

Richard Klein

Univ. Grenoble Alpes

Replication Crisis

Replication Crisis

Theoretical concern

Replication Crisis

Theoretical 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. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

Replication Crisis

Evidence of a problem

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

Replication Crisis

Addressing the problem

Replication Crisis

Addressing the problem

- Many potential causes for replication failures
 - False positives
 - Incompetent replicators
 - Unknown moderators (imprecise theory?)
 - Contextual differences
 - Etc.

Replication Crisis

Addressing the problem

- Many potential causes for replication failures
 - False positives
 - Incompetent replicators
 - Unknown moderators (imprecise theory?)
 - Contextual differences
 - Etc.
- Solution depends on the cause

Replication Crisis

Addressing the problem

- Many potential causes for replication failures
 - False positives
 - Incompetent replicators
 - Unknown moderators (imprecise theory?)
 - Contextual differences
 - Etc.
- Solution depends on the cause
- What should we expect of replications? What does replication "look like"? (statistically, practically)

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.

- 5 "Many Labs" projects completed or in-progress.

Many Labs Projects

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

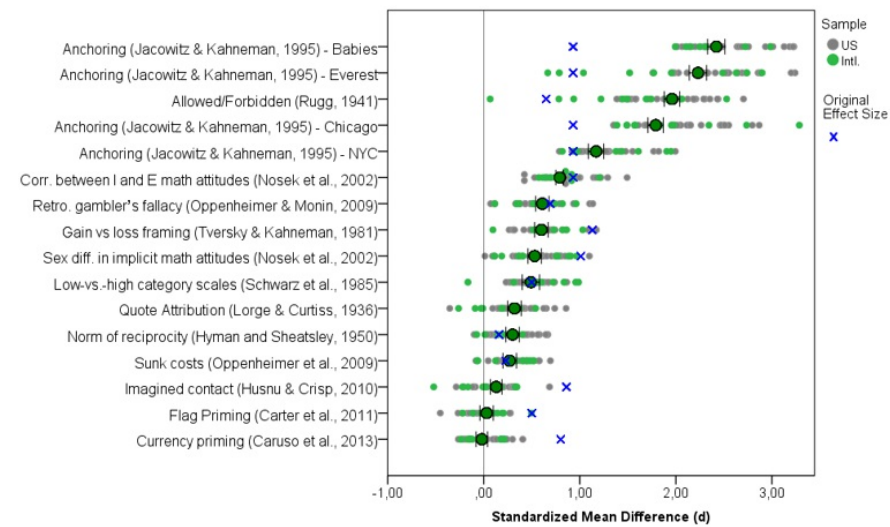
- 5 "Many Labs" projects completed or in-progress.
- I'm presenting Many Labs 2 (December)

Many Labs Projects

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

- 5 "Many Labs" projects completed or in-progress.
- I'm presenting Many Labs 2 (December)
- Same thing as Many Labs 1 (2014), but much bigger.

Many Labs Projects



Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.

Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.
- Replicated 28 studies

Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.
- Replicated 28 studies
 - Selected for impact, diversity of content, possibility for variation across sites (more at osf.io/8cd4r/)

Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.
- Replicated 28 studies
 - Selected for impact, diversity of content, possibility for variation across sites (more at osf.io/8cd4r/)
 - Split across two study "packages" due to length

Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.
- Replicated 28 studies
 - Selected for impact, diversity of content, possibility for variation across sites (more at osf.io/8cd4r/)
 - Split across two study "packages" due to length
 - Computerized in Qualtrics, random presentation order

Many Labs 2

- **Goal:** Replicate studies all around the world, compare variation by context/population.
- Replicated 28 studies
 - Selected for impact, diversity of content, possibility for variation across sites (more at osf.io/8cd4r/)
 - Split across two study "packages" due to length
 - Computerized in Qualtrics, random presentation order
 - Each reviewed by original authors or other experts (Registered Report)

Many Labs 1 Map (2014)



Many Labs 2 Map (2018)

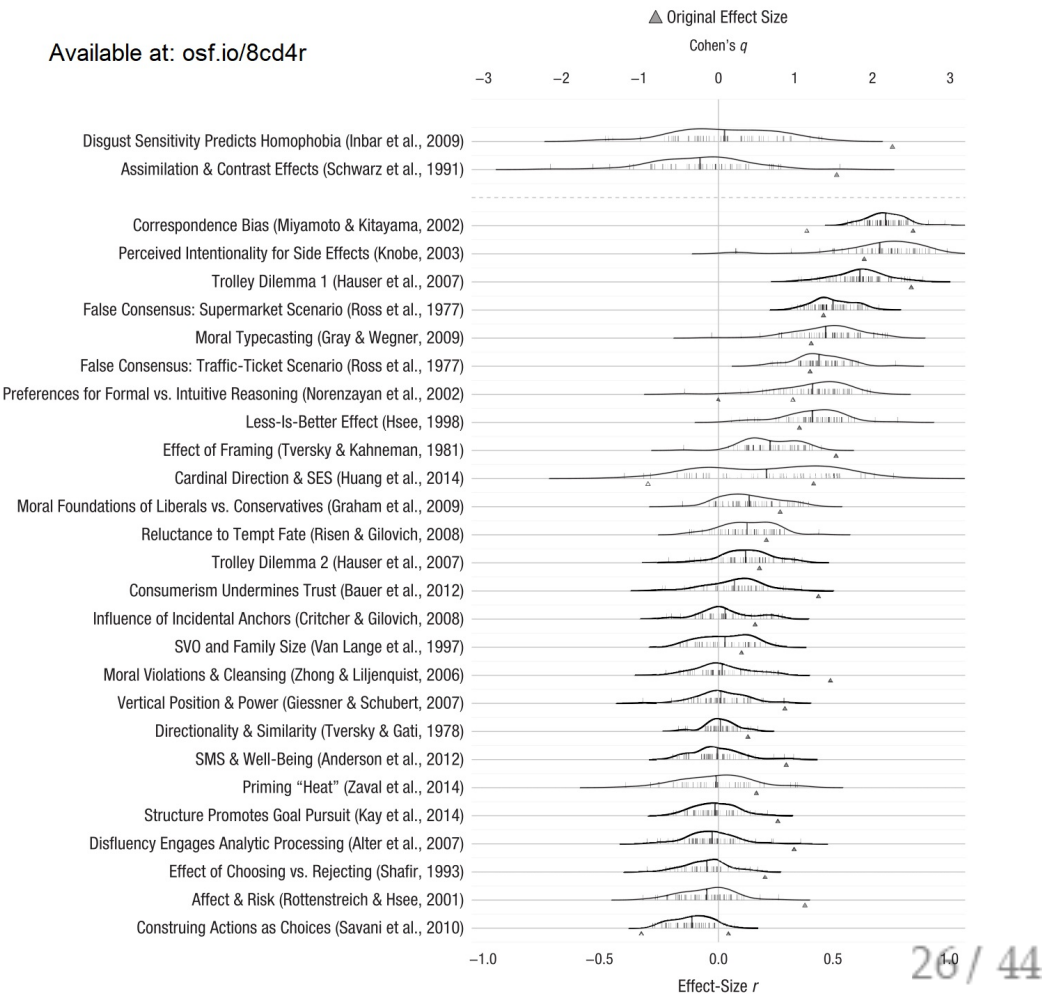


Many Labs 2

- 125 samples
- 36 countries
- 16 languages
- 15,305 participants

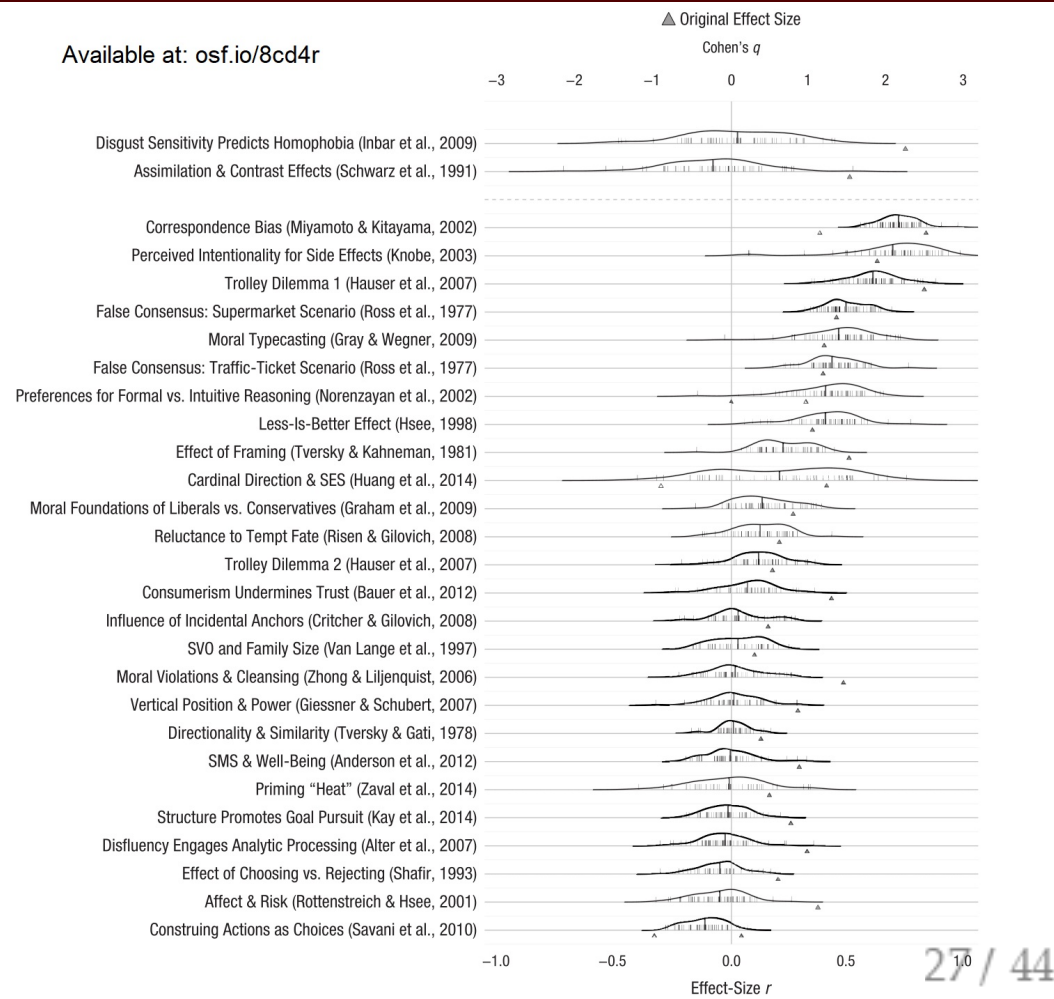


Results



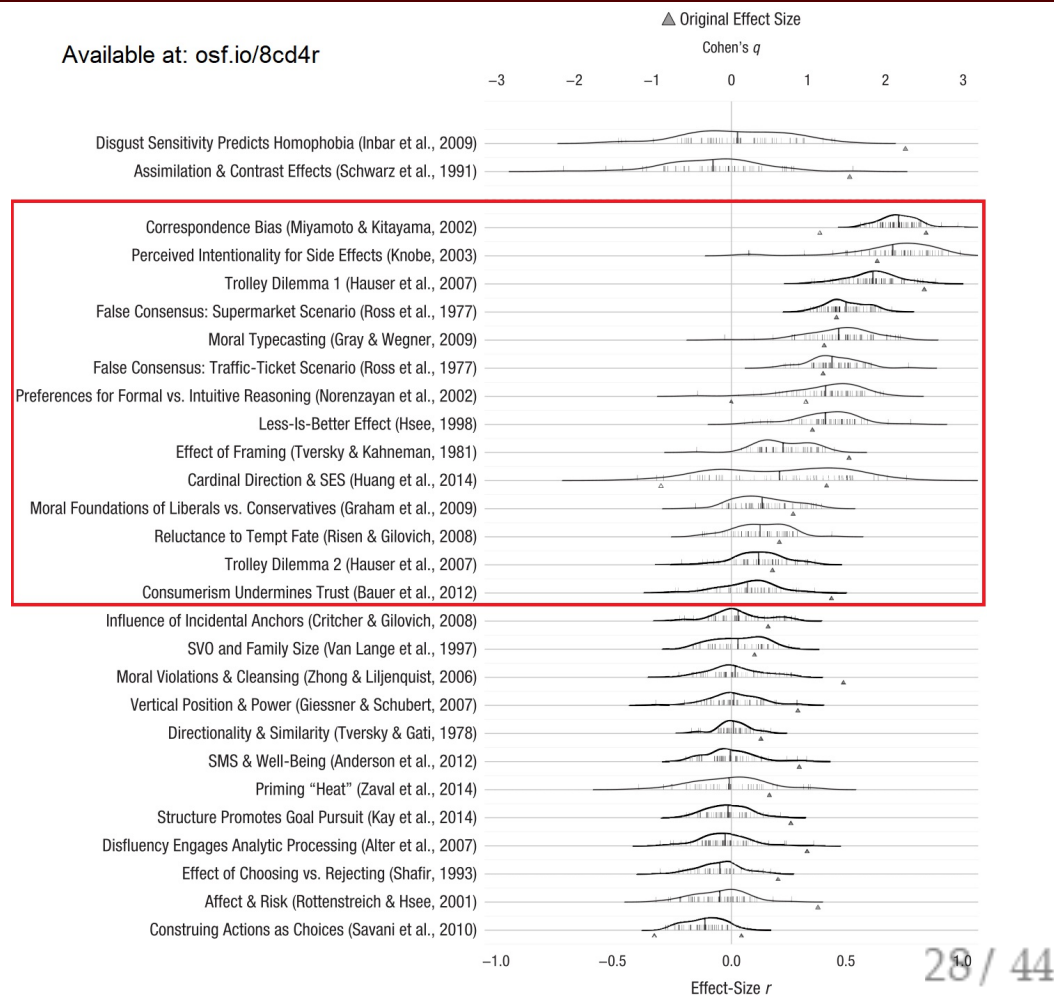
Results

- 14/28 successful



Results

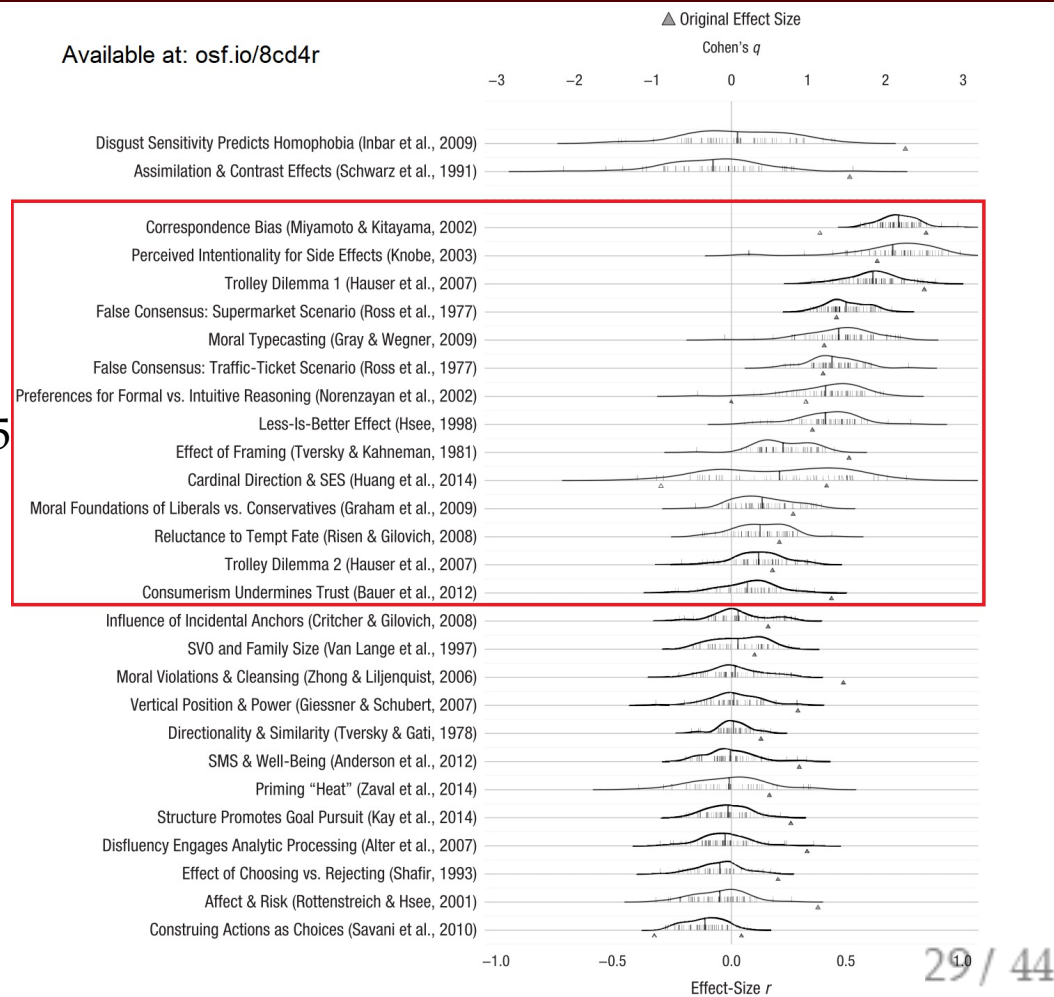
- 14/28 successful



Results

- 14/28 successful
- 21/28 smaller effect
- Med. original $d = 0.60$
- Med. replication $d = 0.15$

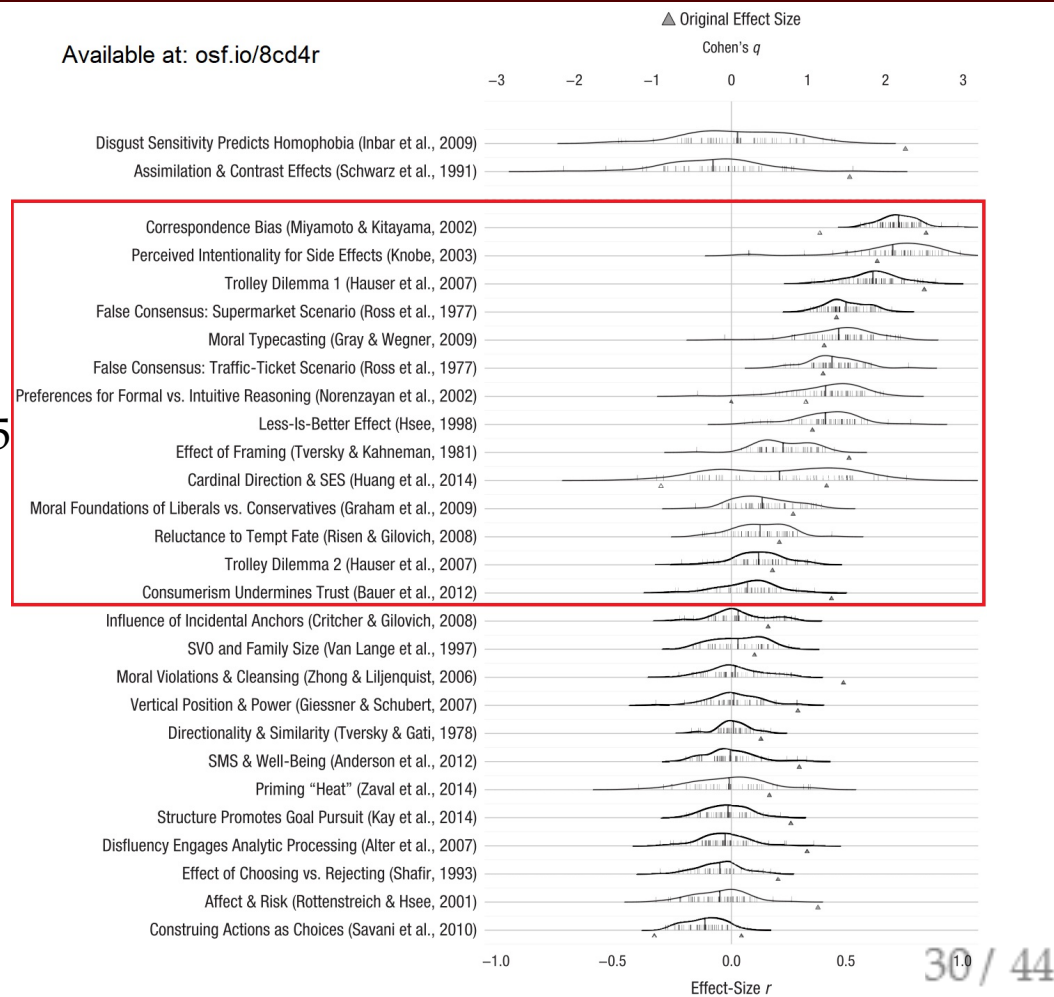
Available at: osf.io/8cd4r



Results

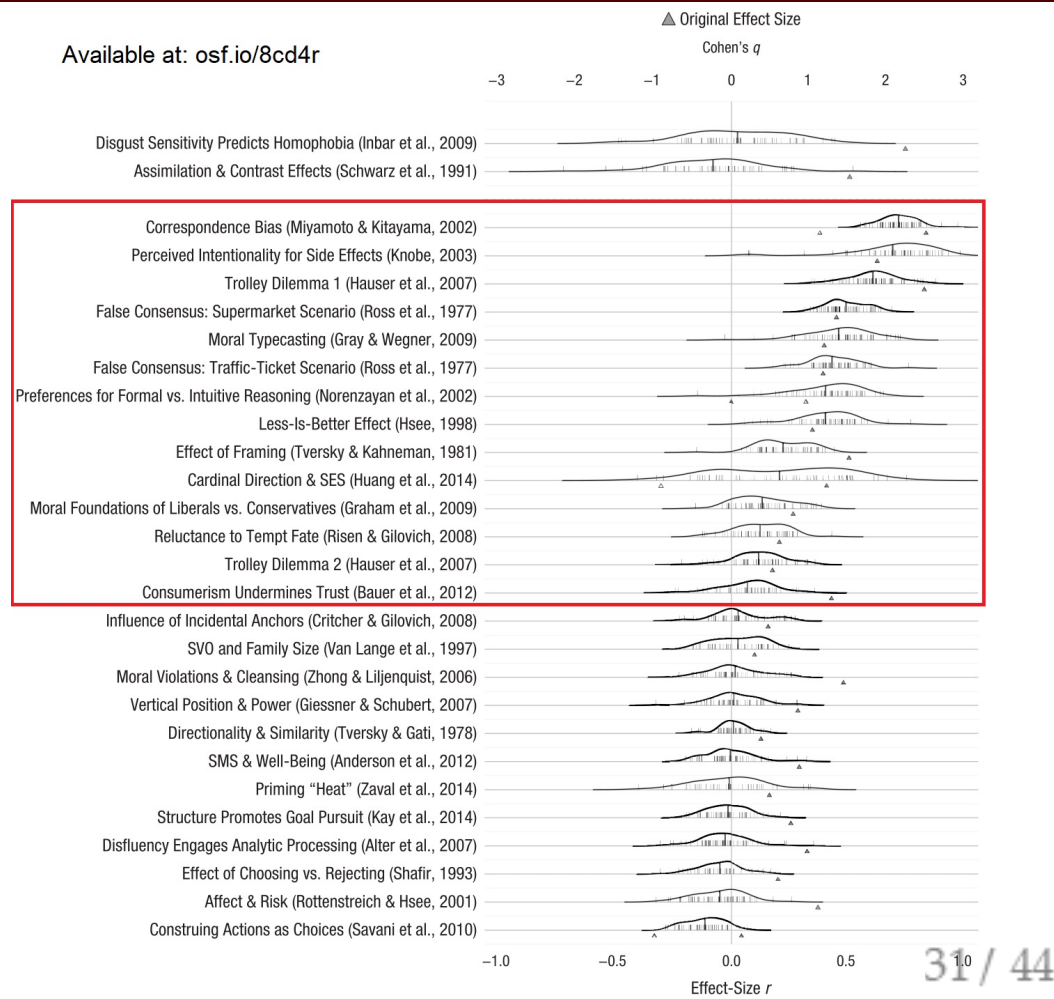
- 14/28 successful
- 21/28 smaller effect
- Med. original $d = 0.60$
- Med. replication $d = 0.15$
- (boring)

Available at: osf.io/8cd4r



Heterogeneity

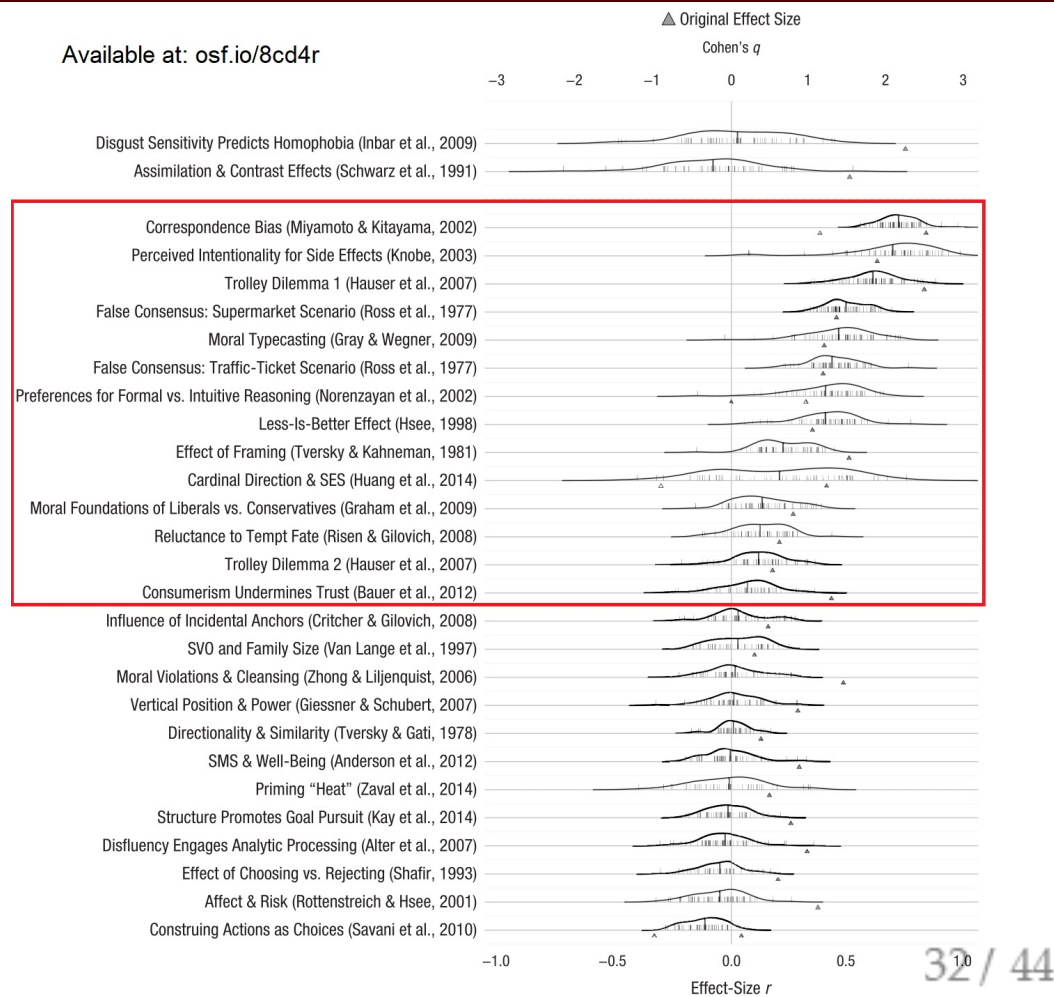
Available at: osf.io/8cd4r



Heterogeneity

- 11/28 $Q < .001$
 - Sig. variability

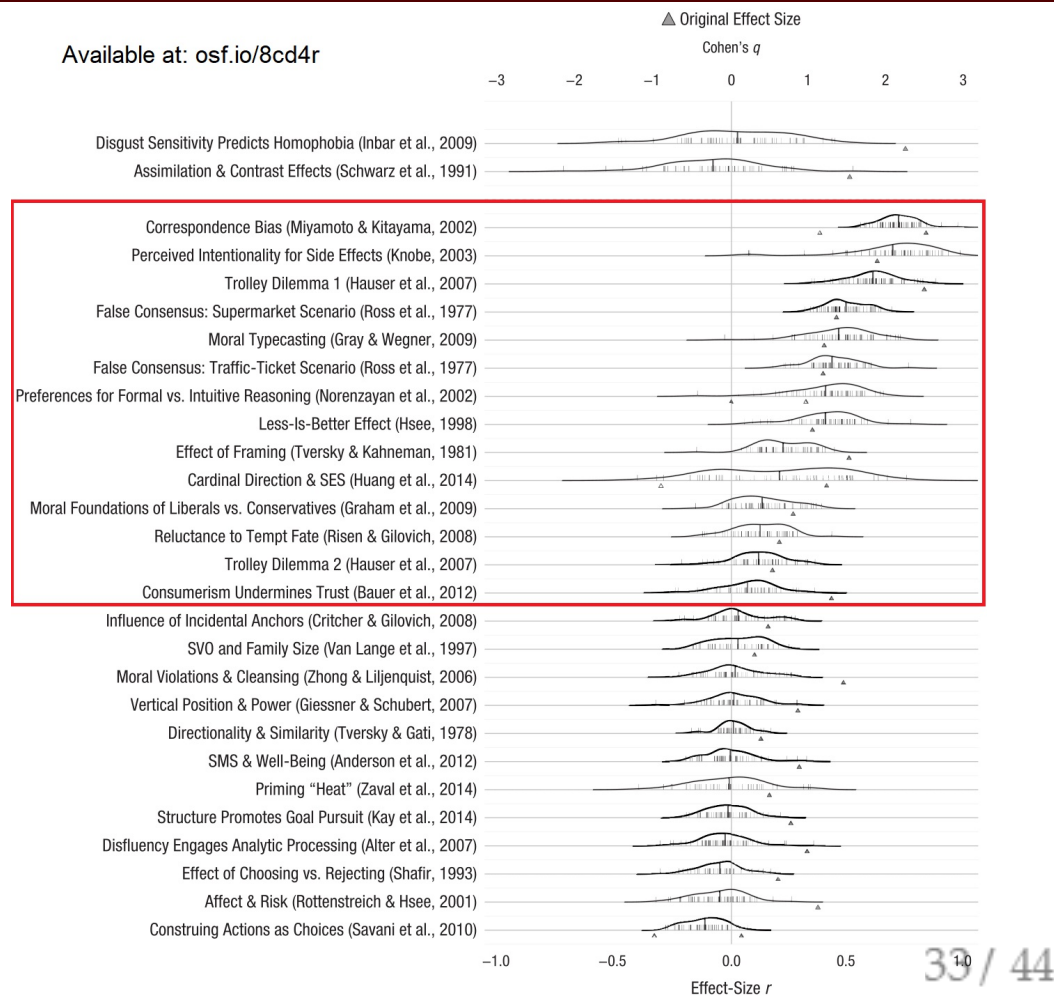
Available at: osf.io/8cd4r



Heterogeneity

- 11/28 $Q < .001$
 - Sig. variability
- HOWEVER:
 - 26/28 $\text{Tau} \leq 0.1$
 - Often 0

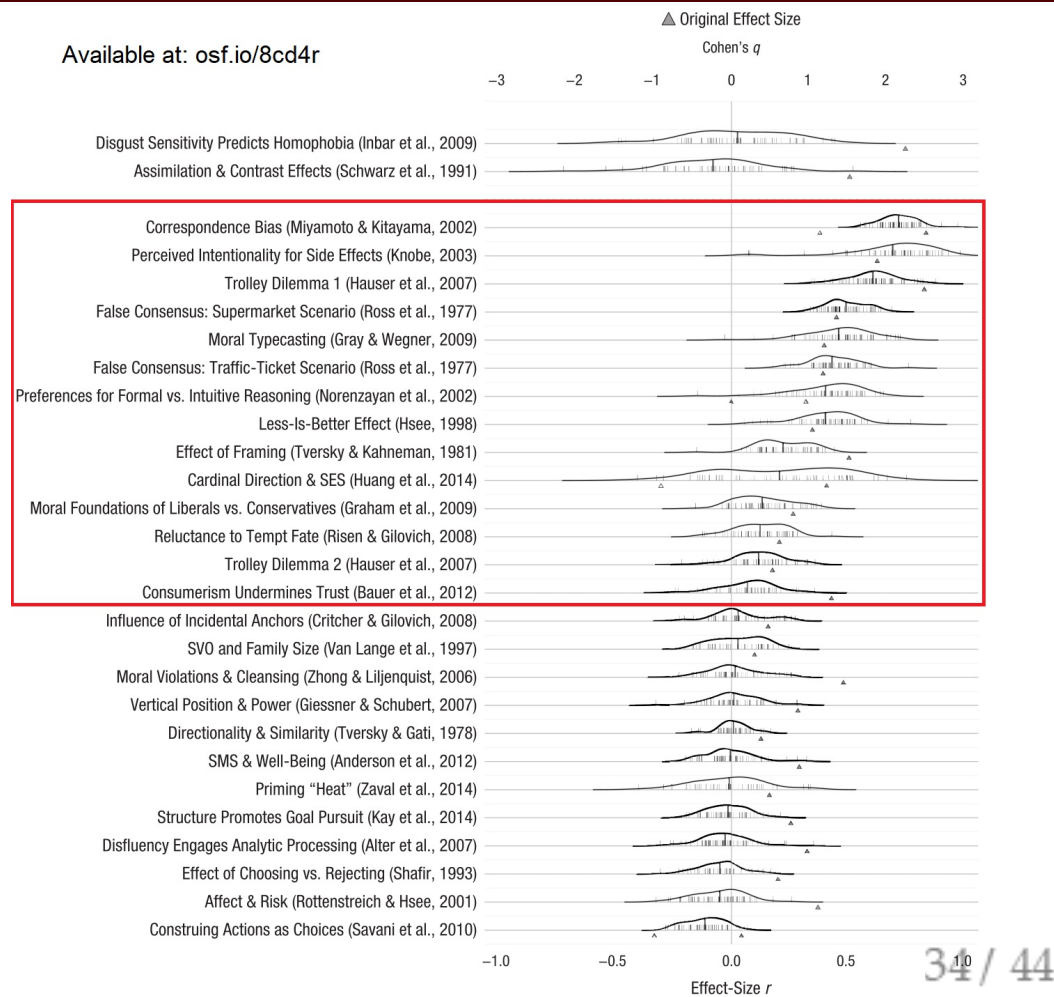
Available at: osf.io/8cd4r



Heterogeneity

- 11/28 $Q < .001$
 - Sig. variability
- HOWEVER:
 - 26/28 $\text{Tau} \leq 0.1$
 - Often 0
- Wait... what?

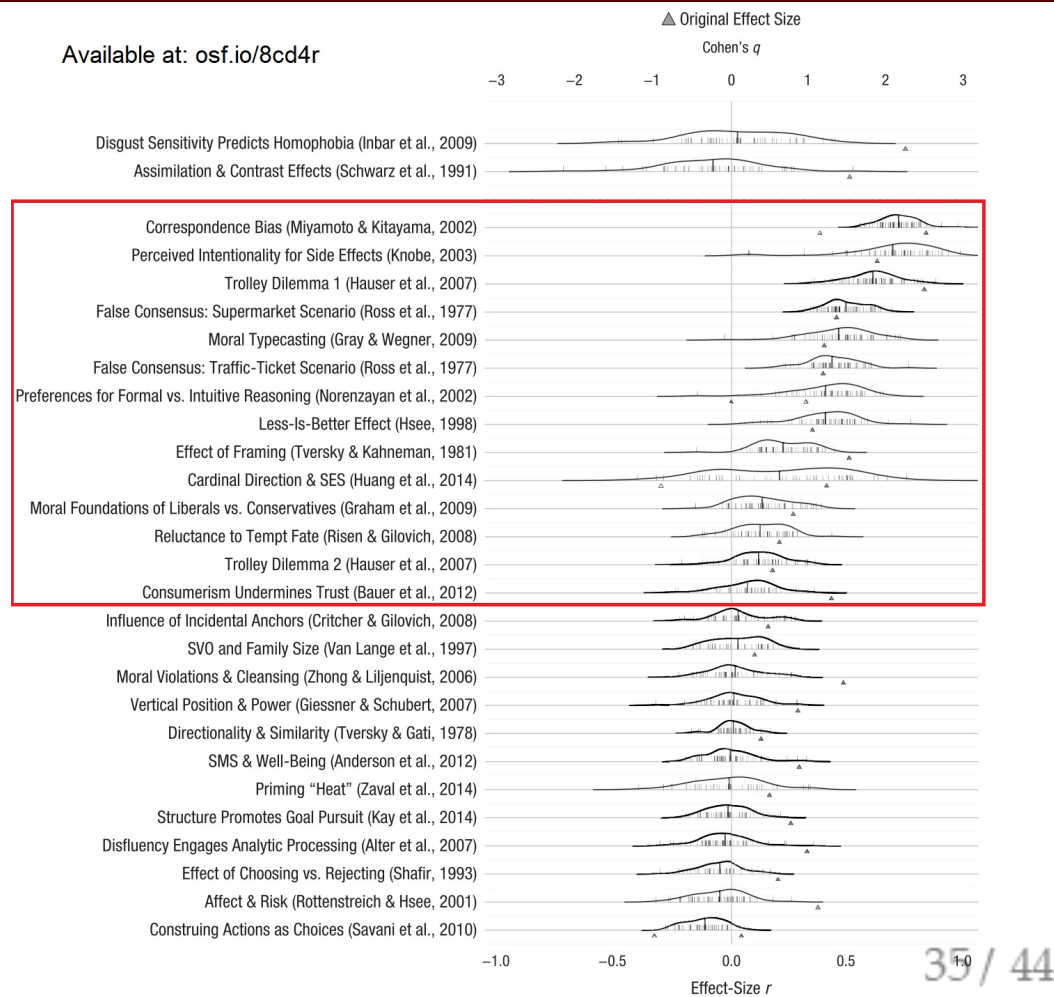
Available at: osf.io/8cd4r



Heterogeneity

- 11/28 $Q < .001$
 - Sig. variability
- HOWEVER:
 - 26/28 $\tau \leq 0.1$
 - Often 0
- Wait... what?
- Mostly sampling error
 - $N = \sim 80$ per site

Available at: osf.io/8cd4r



Discussion

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically
 - Good news for global collaboration

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically
 - Good news for global collaboration
- Furthers understanding of replication
 - Informs power analysis (Kenny & Judd, 2019)

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically
 - Good news for global collaboration
- Furthers understanding of replication
 - Informs power analysis (Kenny & Judd, 2019)
- Replication rate aligns with other projects
 - Is this meaningful?

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically
 - Good news for global collaboration
- Furthers understanding of replication
 - Informs power analysis (Kenny & Judd, 2019)
- Replication rate aligns with other projects
 - Is this meaningful?
- Many studies replicate robustly (and robust replicability is a feasible goal)
 - Failed replications \neq false positive

Discussion

- Low variation across sample/context
 - Despite translation, culture, population differences
 - Not reasonable to assume sample moderators; test empirically
 - Good news for global collaboration
- Furthers understanding of replication
 - Informs power analysis (Kenny & Judd, 2019)
- Replication rate aligns with other projects
 - Is this meaningful?
- Many studies replicate robustly (and robust replicability is a feasible goal)
 - Failed replications \neq false positive
- Open data: <https://osf.io/8cd4r/>
 - CC0, free use

Thanks!

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

Great time to get involved (cos.io/about/news/)

@raklein3
raklein22@gmail.com



financed by
IDEX Université Grenoble Alpes