

Research questions

Problem: Current research paradigms use narrow sets of stimuli that do not fully probe the conceptual robustness of the effect.

- How much variation is there in the study designs and in the estimated results when you let several teams conceptually replicate the same hypothesis?
- Can experts predict which claims are more likely to replicate? Within the same claim, can experts predict which research designs are more likely to lead to a successful replication?

Goals of our work:

- Make transparent how design choices affect research results
- Study by means of a forecasting survey whether the effects of researcher design choices on estimated effect sizes are predictable

Research components

- **5 original hypotheses** in psychology regarding implicit cognition, negotiation and moral judgments
- **Conceptual replications** - 15 research teams *independently* designed studies to test the original hypotheses (main study)
- **Direct replications**
 - Of each original design
 - Of the main study (N > 15,000)
- **Forecasting study** - For each set of materials, independent researchers predicted:
 - Likelihood that a significant result would be found
 - Standardized effect size of the replication

Crowdsourcing Hypothesis Tests

J. Landy, M. Jia, I. Ding, D. Viganola, W. Tierney, A. Dreber, M. Johannesson, T. Pfeiffer, E. Uhlmann (and many others!)

Evidence of large variation in effect sizes across different sets of research designs to test the same hypothesis.

Experts can predict the variation in results across and within hypotheses.

Results of direct and conceptual replications

Hypothesis	Cons. p < .05	Cons. p > .05	Incons. p > .05	Incons. p < .05
Awareness of prejudice	54% (7)	0% (0)	8% (1)	38% (5)
Negotiation and trust	92% (12)	8% (1)	0% (0)	0% (0)
Praise for needless work	77% (10)	8% (1)	8% (1)	8% (1)
Proximal authorities drive legitimacy	42% (5)	25% (3)	25% (3)	8% (1)
Morality and happiness	23% (3)	62% (8)	15% (2)	0% (0)

Direct replications:

- All 5 original designs replicated: the hypotheses were supported when the original materials were used, both in the Main study and in the Replication study

Conceptual replications:

- Fairly consistent support for 2 hypotheses out of 5
- Substantial variability in results across and within hypotheses

Forecasting study: significance and effect size

Fig 1: Histograms of average predictions regarding whether the result is statistically significant ($p < .05$) or not at study-design level.

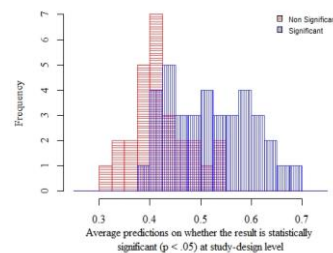
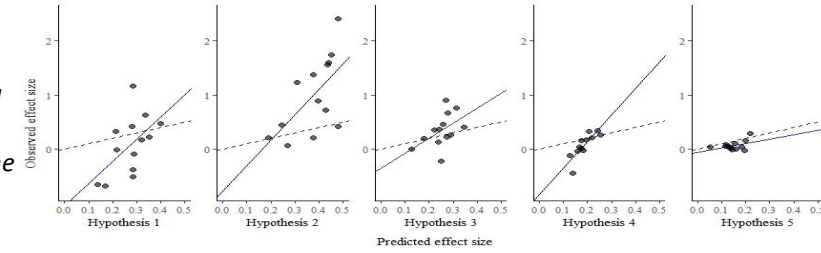
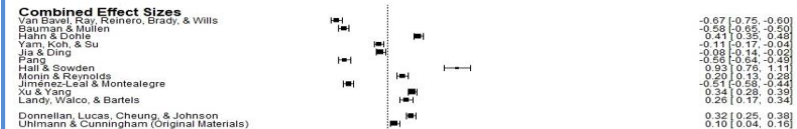


Fig 2: Correlation between average predicted effect size and observed effect size, separately for each of the five hypotheses.

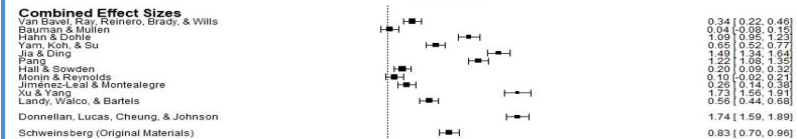


Conceptual replications: forest plots of observed effects

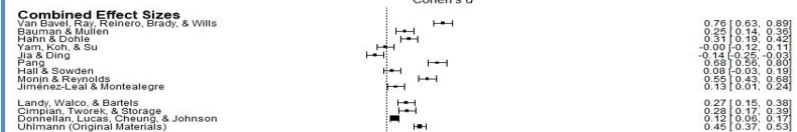
Hypothesis 1



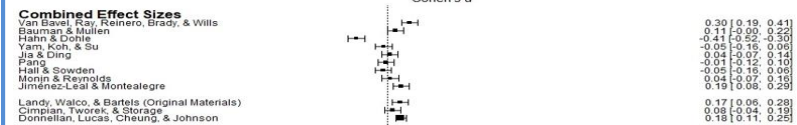
Hypothesis 2



Hypothesis 3



Hypothesis 4



Hypothesis 5

