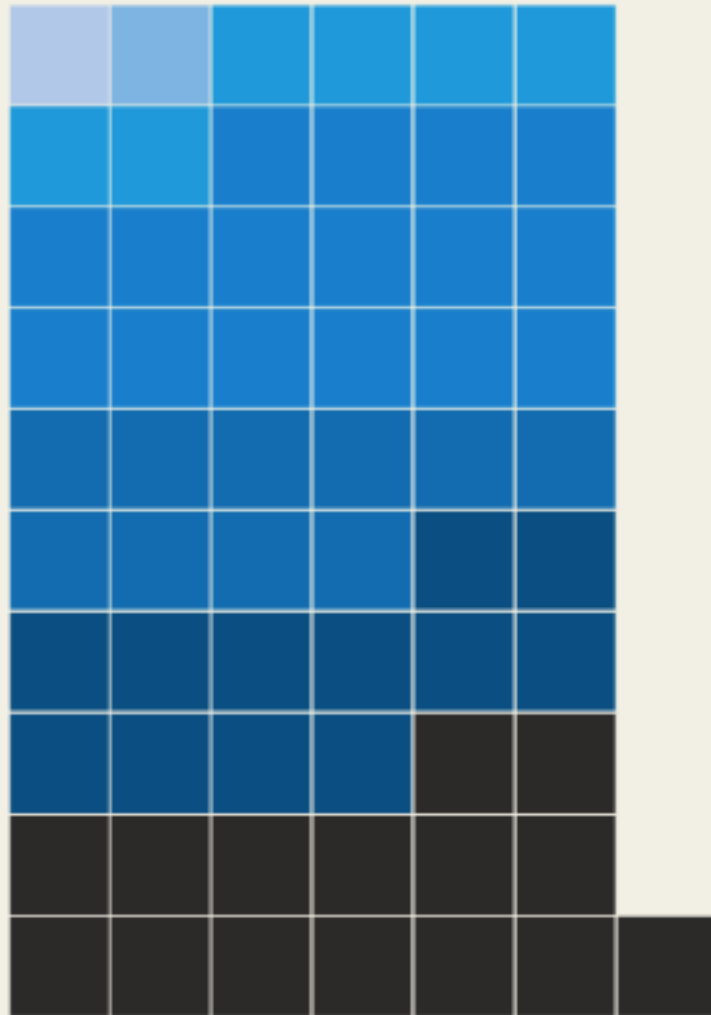


## RELIABILITY TEST

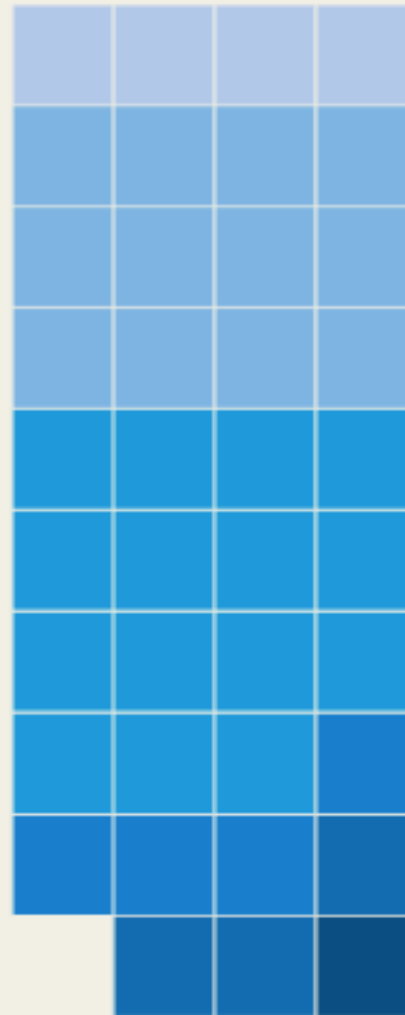
An effort to reproduce 100 psychology findings found that only 39 held up.\* But some of the 61 non-replications reported similar findings to those of their original papers.

Did replicate match original's results?

NO: 61



YES: 39



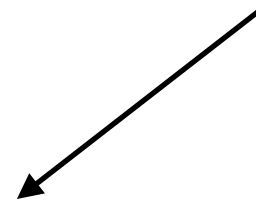
Replicator's opinion: How closely did findings resemble the original study:

- Virtually identical
- Extremely similar
- Very similar
- Moderately similar
- Somewhat similar
- Slightly similar
- Not at all similar

\* based on criteria set at the start of each study

270 authors tried to replicate 100 experiments drawn from high profile Psychology journals - *Psychological Science*, *Journal of Personality and Social Psychology*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

- Button et al. (2013), *Nature Reviews Neuroscience*, small sample size undermines the reliability of neuroscience. Nord et al., (2017), *Journal of Neuroscience*, highlight wide heterogeneity in power in neuroscience studies.



**Table 2. Median, maximum, and minimum power subdivided by study type**

Group	Median power (%)	Minimum power (%)	Maximum power (%)	2.5 <sup>th</sup> and 97.5 <sup>th</sup> percentile (based on raw data)	95% HDI (based on GMMs)	Total N
All studies	23	0.05	1	0.05–1.00	0.00–0.72, 0.80–1.00	730
All studies excluding null	30	0.05	1	0.05–1.00	0.01–0.73, 0.79–1.00	638
Genetic	11	0.05	1	0.05–0.94	0.00–0.44, 0.63–0.93	234
Treatment	20	0.05	1	0.05–1.00	0.00–0.65, 0.91–1.00	145
Psychology	50	0.07	1	0.07–1.00	0.02–0.24, 0.28–1.00	198
Imaging	32	0.11	1	0.11–1.00	0.03–0.54, 0.71–1.00	65
Neurochemistry	47	0.07	1	0.07–1.00	0.02–0.79, 0.92–1.00	50
Miscellaneous	57	0.11	1	0.11–1.00	0.09–1.00	38