

Table 5. Categorizing replication success (bold) or failure based on whether the replication global effect size is in the same direction as the original finding and statistically significant under different conditions

					Replication success criterion						
	Original study sample size	Replication sample size	Replication global effect size		replication sample size, p < .05	replication sample size, p < .0001	Original study sample size, p < .05	2.5x original sample size, p < .05	50/group, p < .05	Minimum sample to detect w/ alpha=.05, power=.80	
Effect				test							
Correspondence Bias (Miyamoto & Kitayama, 2002)	107	7197	1.82	General Linear Model (main effect)	< 1E-10	< 1E-10	4.65E-09	< 1E-10	< 1E-10	12	
Intentional Side Effects (Knobe, 2003)	78	7982	1.75	Welch Two Sample t-test	< 1E-10	< 1E-10	< 1E-10	< 1E-10	< 1E-10	14	
Trolley Dilemma 1 (Hauser et al., 2007)	2646	6842	1.35	Two Sided Fisher's Exact Test	< 1E-10	< 1E-10	< 1E-10	< 1E-10	NA	20	
False Consensus 1 (Ross et al., 1977)	80	7205	1.18	Welch Two Sample t-test	< 1E-10	< 1E-10	6.98E-06	< 1E-10	5.18E-08	26	
Moral Typecasting (Gray & Wegner, 2009)	69	8002	0.95	Welch Two Sample t-test	< 1E-10	< 1E-10	1.86E-04	3.06E-09	6.55E-06	38	
False Consensus 2 (Ross et al., 1977)	80	7827	0.95	Welch Two Sample t-test	< 1E-10	< 1E-10	6.06E-05	2.04E-10	6.64E-06	38	
Intuitive Reasoning (Norenzayan et al. 2002)	157	7396	0.86	Welch Two Sample t-test	< 1E-10	< 1E-10	< 1E-10	< 1E-10	3.92E-05	46	
Less is Better (Hsee, 1998)	83	7646	0.78	Welch Two Sample t-test	< 1E-10	< 1E-10	6.18E-04	5.76E-08	1.69E-04	54	
Framing (Tversky & Kahneman, 1981)	181	7228	0.40	Two Sided Fisher's Exact Test	< 1E-10	< 1E-10	0.031	6.29E-04	NA	200	
Direction & SES (Huang et al., 2014)	180	6591	0.40	Welch Two Sample t-test	< 1E-10	< 1E-10	0.080	5.47E-03	0.0498	200	
Moral Foundations (Graham et al., 2009)	1209	6966	0.29	Fisher r-to-Z test (1 cor)	< 1E-10	< 1E-10	4.12E-07	< 1E-10	0.318	376	
Trolley Dilemma 2 (Hauser et al., 2007)	2612	7923	0.25	Two Sample t-test	< 1E-10	< 1E-10	4.85E-08	< 1E-10	NA	506	
Tempting Fate (Risen & Gilovich, 2008)	120	4599	0.18	Two Sample t-test	< 1E-10	< 1E-10	0.325	0.119	0.369	972	
Priming consumerism (Bauer et al., 2012)	78	6608	0.12	Two Sample t-test	< 1E-10	< 1E-10	0.594	0.399	0.546	2184	
Disgust & Homophobia (Inbar et al., 2009)	42	7041	0.05	Fisher r-to-Z test (2 cor)	0.024	0.024	0.871	0.788	0.794	6283	
Incidental Anchors (Critcher & Gilovich, 2008)	200	6826	0.04	Two Sample t-test	0.092	0.092	0.773	0.649	0.839	19626	
Position & Power (Giessner & Schubert, 2007)	64	7890	0.03	Two Sample t-test	0.162	0.162	0.900	0.842	0.875	34886	
Direction and Similarity (Tversky & Gati, 1978)	154	3549	0.01	One Sample t-test	0.550	0.550	0.973	0.983	0.953	313958	
Moral Cleansing (Zhong & Liljenquist, 2006)	27	7001	0.00	Two Sample t-test	0.910	0.910	0.994	0.991	0.989	NA	
Structure and goal-pursuit (Kay et al., 2014)	67	6506	-0.02	Welch Two Sample t-test	0.347	0.347	0.924	0.880	0.907	NA	
Social Value Orientation (Van Lange et al., 1997)	536	6234	-0.03	Fisher r-to-Z test (1 cor)	0.183	0.183	0.697	0.537	0.908	NA	
Incidental disfluency (Alter et al., 2007)	41	6935	-0.03	Two Sample t-test	0.171	0.171	0.917	0.868	0.870	NA	
Priming warmth (Zaval et al., 2014)	192	4204	-0.03	Two Sample t-test	0.274	0.274	0.816	0.712	0.866	NA	
SMS & Well-Being (Anderson et al., 2012)	116	6905	-0.04	Two Sample t-test	0.079	0.079	0.820	0.719	0.833	NA	
Assimilation and Contrast (Schwarz et al., 1991)	100	7460	-0.07	Fisher r-to-Z test (2 cor)	0.002	0.002	0.734	0.583	0.734	NA	
Affect and Risk (Rottenstreich & Hsee, 2001)	40	7450	-0.08	Two Sided Fisher's Exact Test	0.002	0.002	0.831	0.735	NA	NA	
Choosing or Rejecting (Shafir, 1993)	170	7901	-0.13	One Sample Z-test	5.47E-10	5.47E-10	0.186	0.079	0.314	NA	
				Generalized Linear Mixed Model with binomial (logit) link (main effect)							
Actions are Choices (Savani et al. 2010)	218	5882	-0.18		8.04E-06	8.04E-06	NA	NA	NA	NA	
Successful Replications					15	14	11	12	8		
Unsuccessful Replications					13	14	16	15	16		
Success Rate					54%	50%	41%	44%	33%		

Notes: Findings are ordered by replication global effect size with negative values indicating effects in the opposite direction of the original WEIRD sample. Two effect sizes (Inbar et al., 2009; Schwarz et al., 1991) are on a different metric (Cohen's q) than the rest (Cohen's d). All p-values calculated based on the observed replication global effect size with different assumptions of sample size or alpha criterion (.05 or .0001). Replication success criteria p-values are bold if they meet criteria for successful replication. Replication success criteria p-values are in italics if the replication global effect size was in the opposite direction of the original WEIRD sample. Replication success could not be determined based on the original study sample size, 2.5x original study sample size, and 50/group for Savani et al. (2010) because this information could not be computed for the test used in this effect. Replication success could not be determined based on 50/group if the test was a Two Sided Fisher's Exact Test (4 findings), because this would require making strong assumptions on how the sample size per group is distributed in the 2x2 frequency table. Power analyses are conducted using the Cohen's d and Cohen's q values of the replication effect sizes. Note that if another effect size was used in the original study (e.g. correlation, odds ratio, proportion), these were transformed to Cohen's d values. The last column shows the sample size needed to detect a significant effect in the same direction of the original finding for the observed global effect size with alpha = .05 and power = .80. Cells are marked NA if the global effect size was in the opposite direction of the original finding.