Replicability-Index

Improving the replicability of empirical research

Personalized P-Values for 200+ Social/Personality Psychologists

January 19, 2021 Alpha Wars, False Discovery Rate, Replicability, Replicability Ranking, Science-Wise False Discovery Rate, Social Psychology

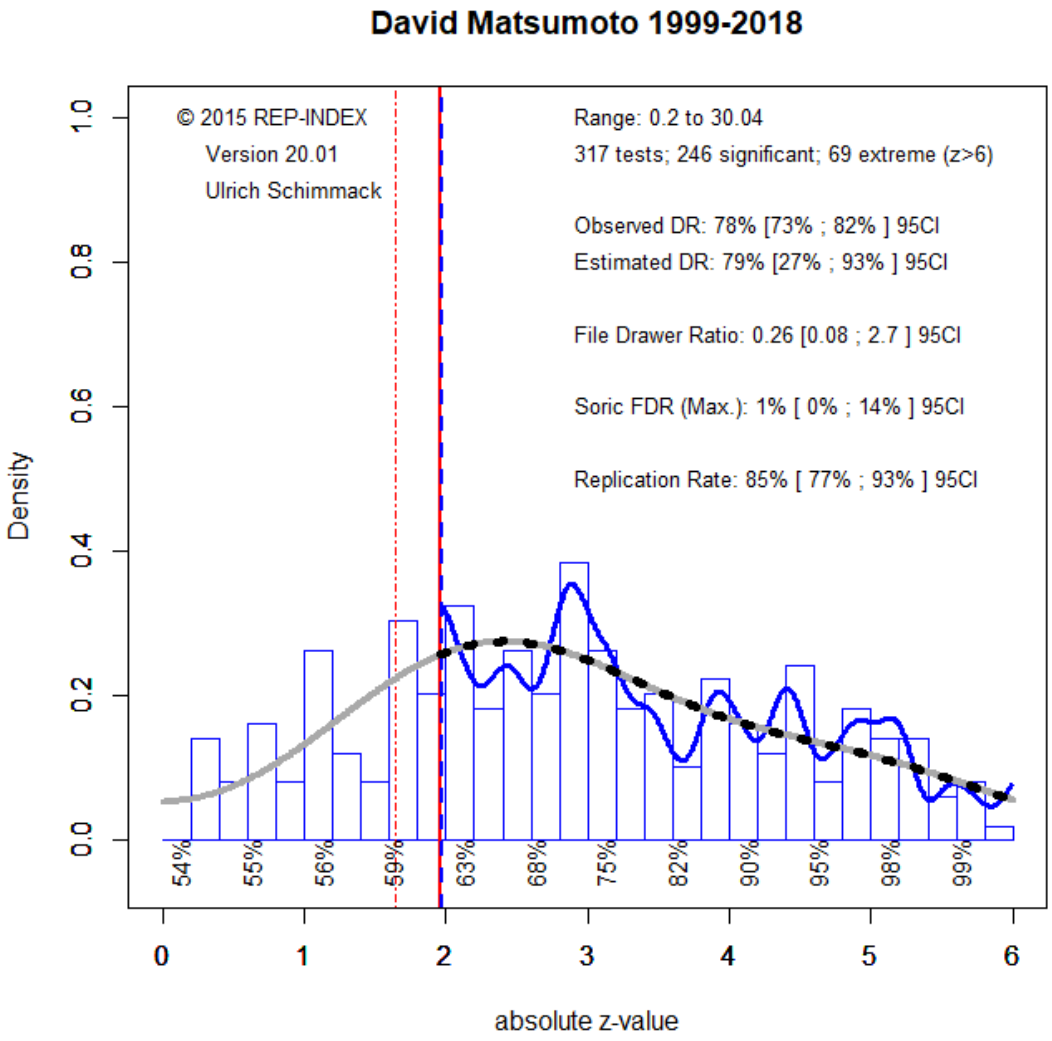
Last update 1/29/2021 (The table will be updated when new information becomes available).

Introduction

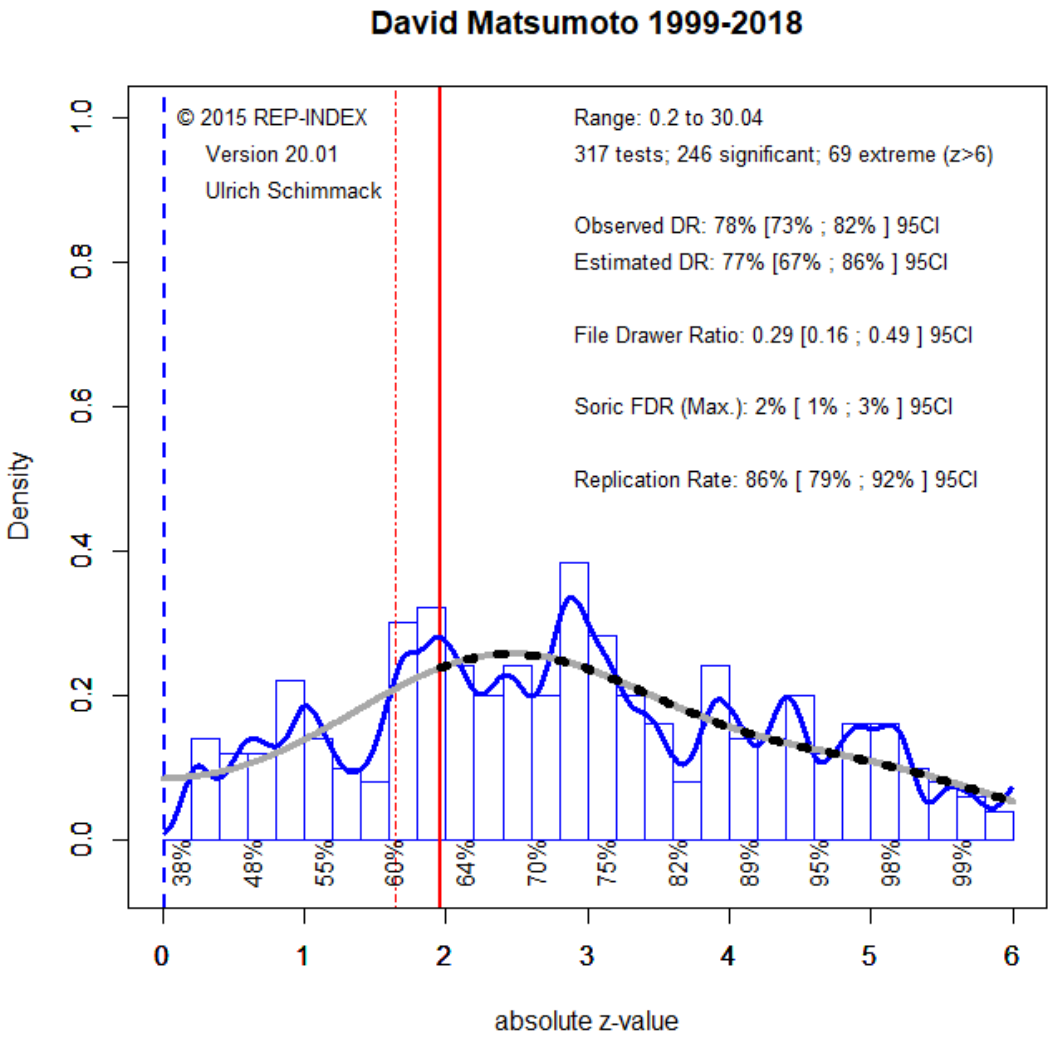
Since Fisher invented null-hypothesis significance testing, researchers have used p < .05 as a statistical criterion to interpret results as discoveries worthwhile of discussion (i.e., the null-hypothesis is false). Once published, these results are often treated as real findings even though alpha does not control the risk of false discoveries.

Statisticians have warned against the exclusive reliance on p < .05, but nearly 100 years after Fisher popularized this approach, it is still the most common way to interpret data. The main reason is that many attempts to improve on this practice have failed. The main problem is that a single statistical result is difficult to interpret. However, when individual results are interpreted in the context of other results, they become more informative. Based on the distribution of p-values it is possible to estimate the maximum false discovery rate (Bartos & Schimmack, 2020; Jager & Leek, 2014). This approach can be applied to the p-values published by individual authors to adjust p-values to keep the risk of false discoveries at a reasonable level, FDR < .05.

Researchers who mainly test true hypotheses with high power have a high discovery rate (many p-values below .05) and a low false discovery rate (FDR < .05). Figure 1 shows an example of a researcher who followed this strategy (for a detailed description of z-curve plots, see Schimmack, 2021).



We see that out of the 317 test-statistics retrieved from his articles, 246 were significant with alpha = .05. This is an observed discovery rate of 78%. We also see that this discovery rate closely matches the estimated discovery rate based on the distribution of the significant p-values, p < .05. The EDR is 79%. With an EDR of 79%, the maximum false discovery rate is only 1%. However, the 95%CI is wide and the lower bound of the CI for the EDR, 27%, allows for 14% false discoveries.



When the ODR matches the EDR, there is no evidence of publication bias. In this case, we can improve the estimates by fitting all p-values, including the non-significant ones. With a tighter CI for the EDR, we see that the 95%CI for the maximum FDR ranges from 1% to 3%. Thus, we can be confident that no more than 5% of the significant results wit alpha = .05 are false discoveries. Readers can therefore continue to use alpha = .05 to look for interesting discoveries in Matsumoto's articles.

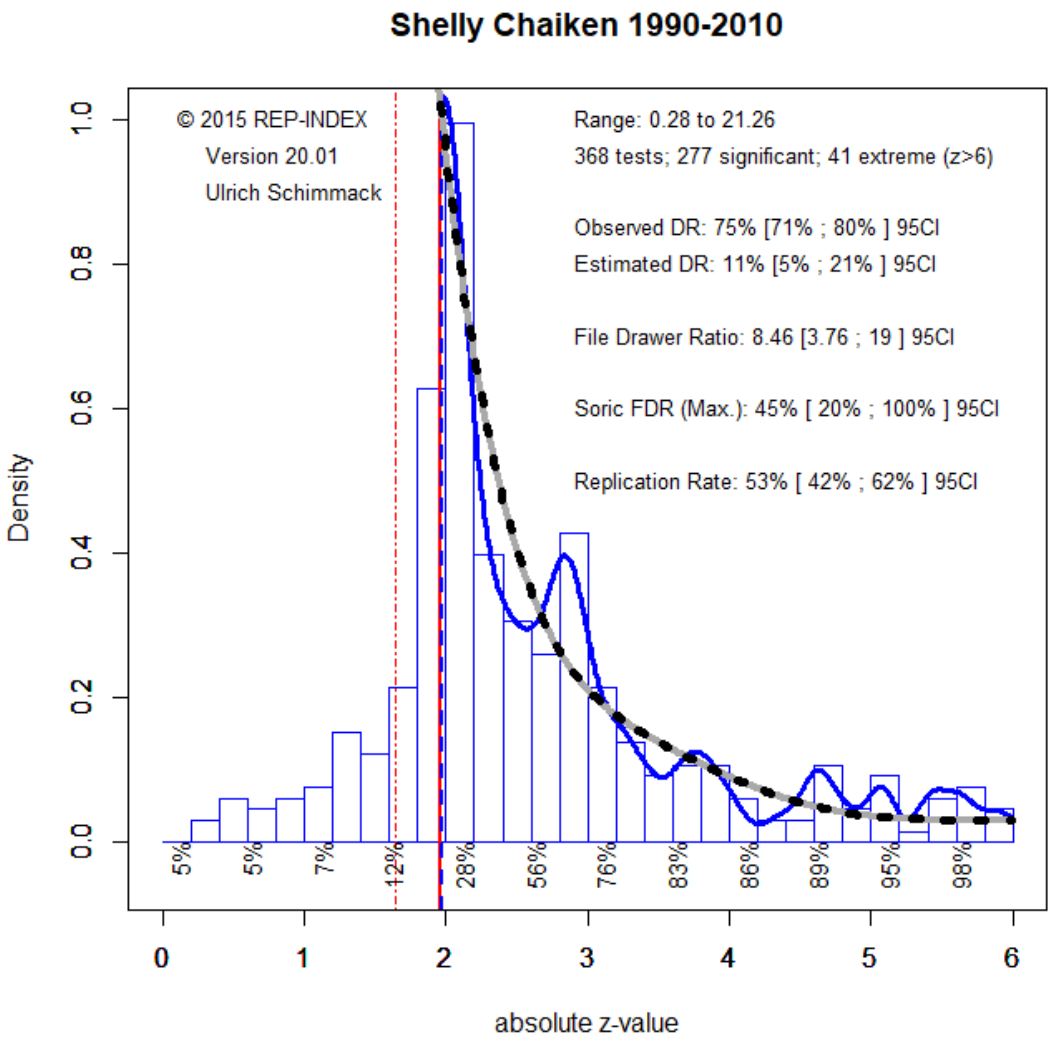
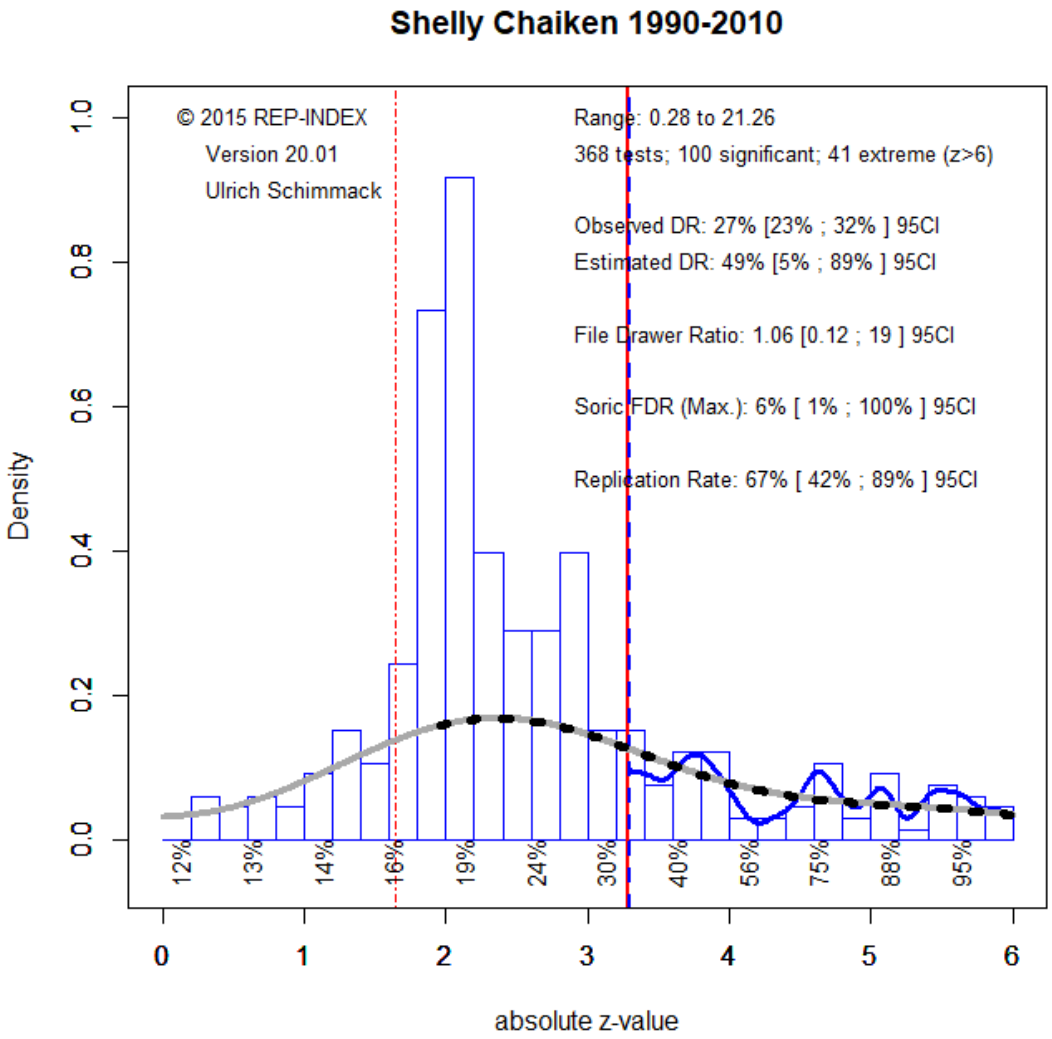


Figure 3 shows the results for a different type of researcher who took a risk and studied weak effect sizes with small samples. This produces many non-significant results that are often not published. The selection for significance inflates the observed discovery rate, but the z-curve plot and the comparison with the EDR shows the influence of publication bias. Here the ODR is similar to Figure 1, but the EDR is only 11%. An EDR of 11% translates into a large maximum false discovery rate of 45%. In addition, the 95%CI of the EDR includes 5%, which means the risk of false positives could be even higher than 45%. In this case, using alpha = .05 to interpret results as discoveries is very risky. Clearly, $p < .05$ means something very different when reading an article by David Matsumoto or Shelly Chaiken.

Rather than dismissing all of Chaiken's results, we can try to lower alpha to reduce the false discovery rate. If we set alpha to .001, most of the just significant results are no longer considered discoveries. Now the EDR is even higher than the ODR because a large pile of just significant results with alpha = .05 were observed, but not predicted by the model. Assuming that p-values below .001 come from a different population of studies, the FDR is now 6% and low enough to warrant inspection of the findings that meet the alpha = .001 threshold. This way 100 of the 277 significant results with p = .05 are still interpretable.



The rankings below are based on automatrically extracted test-statistics from 40 journals (List of journals). The results should be interpreted with caution and treated as preliminary. They depend on the specific set of journals that were searched, the way results are being reported, and many other factors. The data are available (data.drop) and researchers can exclude articles or add articles and run their own analyses using the z-curve package in R (https://replicationindex.com/2020/01/10/z-curve-2-0/).

I am also happy to receive feedback about coding errors. I also recommended to hand-code articles to adjust alpha for focal hypothesis tests. This typically lowers the EDR and increases the FDR. For example, the automated method produced an EDR of 31 for Bargh, whereas hand-coding of focal tests produced an EDR of 12 (Bargh-Audit).

And here are the rankings. The results are fully automated and I was not able to cover up the fact that I placed only #108 out of 221 in the rankings. In another post, I will explain how researchers can move up in the rankings. Of course, one way to move up in the rankings is to increase statistical power. The rankings will be updated in a couple of months when articles from 2020 have been added.

Despite the preliminary nature, I am confident that the results provide valuable information. Until know all p-values below .05 have been treated as if they are equally informative. The rankings here show that this is not the case. While p = .02 can be informative for one researcher, p = .002 may still entail a high false discovery risk for another researcher.

Name	Tests	ODR	EDR	ERR	FDR	ALPHA
Virgil Zeigler-Hill	676	76	80	86	1	.05
David Matsumoto	327	78	79	85	1	.05
Matthew D. Lieberman	377	74	76	80	2	.05
Linda J. Skitka	434	74	73	81	2	.05
Steven J. Heine	566	78	73	80	2	.05
Mahzarin R. Banaji	786	73	72	78	2	.05
David P. Schmitt	241	78	72	80	2	.05
John M. Zelenski	149	69	70	77	2	.05
Kurt Gray	397	76	69	79	2	.05
Phoebe C. Ellsworth	551	76	69	75	2	.05
Richard W. Robins	245	82	68	76	2	.05
Michael E. McCullough	311	71	68	75	2	.05
Jim Sidanius	408	72	67	75	3	.05
Thomas N Bradbury	310	61	66	72	3	.05
Klaus Fiedler	1262	83	66	75	3	.05
James J. Gross	1022	73	64	76	3	.05
Barbara L. Fredrickson	240	84	64	74	3	.05
Joris Lammers	587	72	63	70	3	.05
Paul Rozin	373	78	62	80	3	.05
Margaret S. Clark	386	79	62	74	3	.05
Emily A. Impett	466	74	62	73	3	.05

Edward L. Deci	243	83	61	66	3	.05
Patricia G. Devine	618	72	60	67	3	.05
Alice H. Eagly	299	78	59	73	4	.05
Jean M. Twenge	374	75	58	63	4	.05
Elaine Fox	453	82	57	80	4	.05
Carol D. Ryff	207	86	55	74	4	.05
Richard M. Ryan	917	78	55	71	4	.05
Rainer Banse	396	81	55	77	4	.05
Rainer Reisenzein	213	66	55	71	4	.05
Lee Jussim	246	77	55	73	4	.05
William A. Cunningham	183	68	53	63	5	.01
Mark Schaller	533	71	53	63	5	.05
B. Keith Payne	684	69	51	71	5	.05
Jordan B. Peterson	288	66	51	81	5	.05
Leaf van Boven	500	78	51	68	5	.05
William B. Swann Jr.	963	79	50	79	5	.05
Daniel M. Wegner	593	80	50	64	5	.05
Shinobu Kitayama	820	77	49	73	6	< .001
Agneta H. Fischer	903	78	49	70	5	.05
Igor Grossmann	196	75	49	69	6	< .001
Brian A. Nosek	660	66	49	82	5	.05
Jennifer S. Lerner	166	84	49	66	5	.05
Richard E. Nisbett	290	78	49	71	5	.001

Tessa V. West	564	73	48	57	6	< .001
S. Alexander Haslam	1124	70	48	64	6	.01
Lisa Feldman Barrett	641	69	47	71	6	< .001
Constantine Sedikides	2294	71	47	72	6	.001
Edward P. Lemay	309	89	47	85	6	< .001
Harry T. Reis	596	73	47	65	6	.01
Paul K. Piff	153	78	46	63	6	< .001
Bertram Gawronski	1346	75	46	78	6	.005
Dacher Keltner	1211	75	46	65	6	< .001
Charles M. Judd	1161	79	46	68	6	< .001
Jens B. Asendorpf	213	82	46	72	6	.01
Susan T. Fiske	886	78	45	73	6	< .001
Nicholas O. Rule	886	69	45	73	6	.05
Jan De Houwer	1637	69	45	72	6	.01
Bernadette Park	1039	76	45	65	6	< .001
Dirk Wentura	699	73	44	70	7	.005
Mark J. Brandt	250	73	44	75	7	.001
Hazel Rose Markus	583	76	44	66	7	< .001
Craig A. Anderson	477	75	44	61	7	.01
Philip E. Tetlock	478	81	43	72	7	.001
Norbert Schwarz	1195	74	43	62	7	< .001
Niall Bolger	406	78	43	68	7	< .001

Paula M. Niedenthal	463	72	42	67	7	.001
Tiffany A. Ito	326	80	42	64	7	< .001
Carol S. Dweck	918	72	42	66	7	.001
Michael Inzlicht	517	64	42	58	7	.001
Ursula Hess	709	78	41	74	8	< .001
Stacey Sinclair	322	69	41	56	8	.01
Duane T. Wegener	901	76	41	60	8	< .001
Jessica L. Tracy	482	77	41	75	7	.001
Richard E. Petty	2536	70	41	63	8	< .001
Malte Friese	517	63	41	60	8	.05
Fritz Strack	628	76	41	59	8	.001
Eric D. Knowles	352	70	41	65	8	.001
Christian S. Crandal	359	75	40	58	8	< .001
John T. Cacioppo	329	78	40	67	8	< .001
Mario Mikulincer	868	88	40	63	8	< .001
Nicholas Epley	1206	76	40	71	8	< .001
Tobias Greitemeyer	1680	71	39	67	8	.001
Michael D. Robinson	1357	78	39	66	8	.005
Wendi L. Gardner	766	68	39	63	8	.01
John F. Dovidio	1798	70	38	62	9	< .001
C. Nathan DeWall	1147	74	37	60	9	.005
Eva Walther	478	83	37	68	9	.01
Antony S. R. Manstead	1497	78	37	63	9	< .001

Jerry Suls	380	77	37	68	9	.01
David M. Buss	342	79	37	78	9	< .001
Batja Mesquita	249	78	36	73	10	.001
Thomas Gilovich	1037	81	36	67	10	.001
Kerry Kawakami	465	68	36	57	9	< .001
Samuel L. Gaertner	300	75	36	61	9	< .001
Lorne Campbell	383	74	36	62	9	.005
Caryl E. Rusbult	225	68	35	61	10	< .001
Steven J. Sherman	808	80	35	66	10	.001
Anthony G. Greenwald	398	73	35	83	10	.01
Nalini Ambady	1167	65	35	58	10	< .001
Matthew Feinberg	249	77	35	70	10	.001
Ulrich Schimmack	337	77	35	68	10	< .001
Claude M. Steele	424	78	34	47	10	.01
Jennifer Crocker	424	68	34	61	10	< .001
Dale T. Miller	441	74	34	66	10	.005
Azim F. Sharif	155	79	34	73	10	.01
Marcel Zeelenberg	776	80	34	83	10	.001
Jeffry A. Simpson	657	77	33	58	11	< .001
Nilanjana Dasgupta	406	75	33	56	10	.005
Daphna Oyserman	474	54	33	53	11	< .001
Russell H. Fazio	995	69	33	59	11	.01

Jennifer A. Richeson	801	68	33	50	11	< .001
Shigehiro Oishi	914	67	33	63	11	.001
Emily Balcetis	488	72	33	66	11	< .001
Karl Christoph Klauer	589	72	32	70	11	.01
Kathleen D. Vohs	908	69	32	53	11	< .001
Russell Spears	1968	76	32	55	11	< .001
Ap Dijksterhuis	672	71	32	53	11	< .001
Arthur Aron	297	67	31	59	12	< .001
Sander L. Koole	722	69	31	55	12	.01
John A. Bargh	556	69	31	54	12	< .001
Ara Norenzayan	198	75	30	59	12	.005
Mark Snyder	546	73	30	64	12	.01
Joshua Aronson	176	79	30	48	12	< .001
Wendy Wood	487	76	30	60	12	.001
Roger Giner-Sorolla	388	77	30	73	12	.001
Arron C. Kay	1145	71	30	49	12	.001
Joel Cooper	265	77	29	44	13	< .001
Klaus R. Scherer	465	84	29	82	13	.01
Michael Harris Bond	324	75	29	85	13	< .001
Yoav Bar-Anan	481	74	29	79	13	< .001
Roy F. Baumeister	2133	71	29	53	13	.01
Adam D. Galinsky	1803	73	28	50	13	.01
Galen V. Bodenhausen	580	74	28	63	14	< .001

Gordon B. Moskowitz	353	73	28	59	13	< .001
Grainne M. Fitzsimons	559	69	28	49	13	.01
Shelly L. Gable	321	64	28	48	13	< .001
Ronald S. Friedman	191	78	28	46	14	.005
Richard J. Davidson	410	65	27	52	14	.01
Kennon M. Sheldon	636	78	27	65	14	.001
Jeff Greenberg	1387	79	27	57	14	.005
Sonja Lyubomirsky	436	76	27	61	14	< .001
Lauren J. Human	238	40	27	43	14	.01
Eliot R. Smith	433	79	26	68	15	< .001
Eli J. Finkel	1237	65	26	56	15	< .001
John T. Jost	687	75	26	64	15	.001
Tom Pyszczynski	984	71	26	58	15	< .001
Jonathan Haidt	353	76	25	75	16	< .001
Elizabeth W. Dunn	341	73	25	60	16	< .001
Felicia Pratto	330	75	25	76	16	< .001
Phillip R. Shaver	508	82	25	73	16	.001
Brent. W. Roberts	156	76	25	66	16	< .001
Roland Neumann	206	78	25	62	16	< .001
Nyla R. Branscombe	1243	72	25	66	16	.001
David A. Pizarro	205	73	24	70	17	.005
Amy J. C. Cuddy	156	82	24	77	17	< .001

Tanya L. Chartrand	416	66	24	39	17	.001
Joseph P. Forgas	907	85	24	60	17	< .001
Paul Bloom	244	83	24	75	17	.01
Mark P. Zanna	663	71	24	49	17	.001
Yoel Inbar	260	69	23	73	18	< .001
Klaus Rothermund	665	76	23	77	18	.01
Peter M. Gollwitzer	1086	68	23	57	18	< .001
Robert S. Wyer	837	82	22	62	19	.001
Laurie A. Rudman	445	78	22	69	18	.01
Michael Ross	1100	73	22	64	18	< .001
Dieter Frey	1493	69	22	58	18	.001
Gabriele Oettingen	727	63	22	48	18	.01
Ed Diener	418	74	22	74	19	.01
Gerald L. Clore	443	72	21	49	20	< .001
Roland Deutsch	331	80	21	76	19	< .001
Andrew J. Elliot	843	83	21	64	19	.01
Wendy Berry Mendes	929	69	21	45	19	< .001
Eddie Harmon-Jones	806	69	21	69	20	.005
Sandra L. Murray	560	67	21	64	19	.001
Robert B. Cialdini	351	75	20	60	21	< .001
Frank D. Fincham	631	74	20	58	21	< .001
James K. McNulty	907	61	20	66	21	< .001
Toni Schmader	530	66	20	61	21	< .001

Benoit Monin	578	71	20	57	21	.001
Wilhelm Hofmann	603	66	20	68	21	.001
Gun R. Semin	163	71	19	70	22	< .001
Boris Egloff	234	83	19	61	22	.01
Marilynn B. Brewer	345	75	19	60	22	.01
Thomas Mussweiler	601	70	19	46	22	< .001
Michael W. Kraus	486	72	18	52	25	< .001
E. Tory. Higgins	1797	71	18	54	24	< .001
David Dunning	689	76	18	69	25	.005
Brandon J. Schmeichel	635	68	18	49	24	.001
Ziva Kunda	216	68	17	61	26	.01
Charles S. Carver	170	80	17	66	25	.01
Steven W. Gangestad	246	56	17	46	25	.01
Simone Schnall	262	61	17	31	25	.01
Jeffrey W Sherman	512	73	17	68	25	< .001
Dolores Albarracin	423	68	16	58	27	.01
Laura A. King	362	76	16	69	27	< .001
Nira Liberman	1129	77	16	67	29	< .001
Lee Ross	359	81	16	65	29	.01
Brad J. Bushman	776	76	16	60	27	.01
Carey K. Morewedge	577	76	16	66	29	< .001
Travis Proulx	171	65	16	64	28	.001

Arie W. Kruglanski	1156	79	16	57	28	.005
Paul W. Eastwick	470	70	15	63	29	< .001
Daniel T. Gilbert	655	67	15	64	29	.001
Steven J. Spencer	554	68	15	40	31	< .001
Nathaniel M Lambert	408	70	15	58	30	< .001
Timothy D. Wilson	669	67	15	64	30	.005
Leandre R. Fabrigar	502	69	14	64	32	.01
Yaacov Trope	1261	75	14	60	32	< .001
Shelley E. Taylor	394	76	14	59	33	.01
William von Hippel	374	66	14	47	34	.005
Dov Cohen	619	71	13	49	34	.001
Mark Muraven	465	64	13	53	34	.001
Oscar Ybarra	288	70	12	58	38	.001
Michael A. Olson	285	70	12	61	37	.005
Gregory M. Walton	437	70	12	41	38	.01
Daniel M. Oppenheimer	167	78	12	60	37	.001
Hans Ijzerman	213	57	10	52	46	< .001
Shelly Chaiken	347	76	10	52	45	< .001

2 bloggers like this.

13 thoughts on "Personalized P-Values for 200+ Social/Personality Psychologists"

Ursula Hess

January 20, 2021 at 6:39 am

Only 801 of the listed 1260 effects were actually taken from research that I was involved in (some seem to stem from articles for which I was editor, others are a mystery to me). On the other hand, the majority of my research is missing. It seems preferable to publish data that is actually based on a more or less representative sample of research actually done by the person with whom that data is associated.

★ Like

★ Ulrich Schimmack

January 20, 2021 at 7:35 am

Thank you for the comment. They are valuable to improve the informativeness of the z-curve analyses.
1. only social/personalty journals and general journals like Psych Science were used (I posted a list of the journals).
I will make clear which journals were used.
2. I am trying to screen out mentions of names as editor, but the program is not perfect. I will look into this and update according.
3. I found a way to screen out more articles where your name appeared in footnotes (thank you).
4. I updated the results and they did improve.
5. Please check the new results.

★ Like

Ursula Hess

January 20, 2021 at 10:37 am

Thank you for the quick response. Some of my research is published in psychophysiology or cognitive journals hence I now understand why so much is missing.

★ Like

★ Ulrich Schimmack

January 20, 2021 at 11:35 am

I figure that research practices can vary once physiological measures are taken or in cognitive studies with within-subject designs. I will eventually do similar posts for other areas.

★ Like

Bill von Hippel

January 23, 2021 at 12:26 am

I'm dismayed (and aghast) to see that I'm almost at the bottom of this list. Any advice on how to investigate this further to see where the problem lies?

★ Like

★ Ulrich Schimmack

January 23, 2021 at 8:51 am

Thank you for your comment.
You can download a file called "William von Hippel-rindex.csv"
It contains all the articles that were used and computes the R-Index based on the z-scores found for that article. The R-Index is a simple way to estimate replicability that works for small sets of test statistics. An R-Index of 50 would suggest that the replicability is about 50%. The EDR would be lower, but is hard to estimate with a small set of test statistics. The file is sorted by the R-Index. Articles with an R-Index below 50 are probably not robust. This is a good way to start diagnosing the problem.

★ Like

Bill von Hippel

January 23, 2021 at 4:51 pm

Hi Uli, that's very helpful – thanks!

But now I'm confused. To start with the worst offenders on my list, I have four papers with an R-Index of 0. I can't tell what two of them are, as your identifier doesn't include the article title or authors, but two of them are clear. The first of those two has large samples, reports a wide variety of large and small correlations, and strikes me as highly replicable. Indeed, study 2 (N=466) is a direct replication of study 1 (N=196) with an even larger

sample. Study 3 goes in a slightly different direction, but mostly relies on the data from Study 2. The other paper reports large samples (Ns = 200) but small effects. We submitted it with only one study, the editor asked for replication, we ran a direct replication with the same sample size and found the same effect. Those are both in the paper. Since then we've tried to replicate it once and have succeeded (that finding isn't yet published).

That's the first issue, and strikes me as the most important. Secondarily, there are at least four or five papers in this list that aren't my own – perhaps more but it's hard to tell what some of the papers are – and the resultant list of papers is only about 1/3 of my empirical publications. Thus, setting aside the most important issue above, I don't have a clear sense of what my actual replicability score would look like with all of my papers.

All the best, Bill

★ Like

★ Ulrich Schimmack

January 23, 2021 at 5:19 pm

please check the number of results. Many papers with R-Index of 0 have only 1 result which is often just a missing value, meaning no results were found. So, you can ignore those.

★ Liked by 1 person

★ Ulrich Schimmack

January 23, 2021 at 5:20 pm

I also made clear which journals were searched for these articles. Please see the list on the blog post.

★ Liked by 1 person

★ Ulrich Schimmack

January 23, 2021 at 5:21 pm

I would also be happy to run an analysis on all of your articles, if you send me the pdfs.

★ Liked by 1 person

Pingback: Jens Forster and the Credibility Crisis in Social Psychology | Replicability-Index

Bill von Hippel

January 23, 2021 at 5:26 pm

There are numerous correlations reported in both papers, along with various mediational analyses in one of them, so definitely not a single result.

With regard to the second issue, the file lists the journal title and year, but that's it. Sometimes I haven't published in that journal in that year, so I know it's not me. Sometimes I have, but in this particular case the only paper I published in that journal in that year has another one of the R = 0 examples, but includes a sample in the millions and a multiverse analysis. There's no chance that could have a replicability index of 0.

★ Liked by 1 person

Bill von Hippel

January 23, 2021 at 5:29 pm

Thanks Uli, very kind of you to offer to run the analysis for me. I've created a dropbox folder with all of my empirical articles in it and shared it with you. Let me know if that doesn't come through. Best, Bill

★ Liked by 1 person