

Running Head: Commentary on Vaidis et al. (2024)

Noise vs. signal:

What can we conclude when a classic finding fails to replicate?

Wilson Cyrus-Lai

STANFORD UNIVERSITY

Warren Tierney

INSEAD

Eric Luis Uhlmann

INSEAD

(Commentary on Vaidis et al., “A Multilab Replication of the Induced-Compliance Paradigm of Cognitive Dissonance”)

ABSTRACT: 164 words

MAIN TEXT: 1845 words

REFERENCES: 1576 words

ENTIRE TEXT: 3715 words

Authors' Note: Please address correspondence to Wilson Cyrus-Lai ([cyruslai@stanford.edu](mailto:cyruslai@stanford.edu))

CONTACT INFORMATION:

Wilson Cyrus-Lai  
Organisational Behaviour Area  
Stanford University  
655 Knight Way, Stanford, CA 94305  
United States of America  
Phone: 650 9000010  
E-mail: cyruslai@stanford.edu

Warren Tierney  
Organisational Behaviour Area / Marketing Area  
INSEAD  
1 Ayer Rajah Avenue 138676  
Singapore  
Phone: 353 87329150  
E-mail: warren.tierney@insead.edu

Eric Luis Uhlmann  
Organisational Behaviour Area  
INSEAD  
1 Ayer Rajah Avenue 138676  
Singapore  
Phone: 65 8468 5671  
E-mail: eric.luis.uhlmann@gmail.com

## Abstract

Vaidis et al. (2024) report a stunning crowdsourced failure to replicate a cornerstone effect in social psychology, specifically dissonance in the induced compliance paradigm. Going beyond most prior replication initiatives, they included careful process measures, showing their experimental manipulation affected perceived choice but that this did not in turn instigate attitude change. Had Vaidis et al. (2024) not observed any effects on a perceived sense of choice, would this render their results uninformative regarding the theory of cognitive dissonance? We think not. Many replicators are *statistical skeptics* who regard implausibly large effect sizes from small underpowered studies as mainly noise rather than signal, rather than *hypothesis skeptics* seeking to falsify the original theoretical claim. From a statistical skeptic's point of view, the originally reported effects on process measures are potential false positives as well. A replication approach that relies on large multisite samples like that of Vaidis et al. (2024) can better distinguish between noise and signal in psychological research than traditional small-scale science.

Cognitive dissonance in the induced compliance paradigm (Croyl & Cooper, 1983; Festinger & Carlsmith, 1959) represents one of the foundational theories and experimental paradigms of social psychology. And yet despite a truly heroic effort, Vaidis et al. were unable to obtain similar results across a host of partner laboratories based in over a dozen nations that recruited a total of over four thousand research participants. The original Croyl and Cooper (1983) research has been widely cited and influential, and was authored by eminent researchers at prestigious academic institutions. And yet it drew a small sample from a single population and was carried out before it became commonplace to limit researcher degrees of freedom (Simmons, Nelson, & Simonsohn, 2011) by committing to one's planned methods and analyses in advance (van't Veer & Giner-Sorolla, 2016; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). As with most crowd initiatives, Vaidis et al. (2024) assembled a coalition of the willing, including researchers whom varied greatly in academic seniority and topic specialization and were based at many different institutions around the world. A familiar dilemma repeats itself, but this with time with much greater stakes given the extraordinary impact and importance of the original work. What should we conclude from a systematic failed replication?

### *The context sensitivity defence*

Initially, the main argument against drawing strong inferences from null results was the likely sensitivity of social judgments and behaviors to hidden moderators such as cultural and population differences (Bargh, 2012a; Gilbert, King, Pettigrew, & Wilson, 2016; Schnall, 2014; Schwarz & Strack, 2014; Stroebe & Strack, 2014; Van Bavel et al., 2016). Effects that are confirmed meta-analytically aggregating across sites do exhibit statistically substantial heterogeneity across populations, with large effect sizes in some samples and small or even

near-zero estimates in others (Klein et al., 2018; Krefeld-Schwalb et al., 2024). However, we believe this variability is at times overstated in that it only relatively rarely involves qualitative differences in effects.

Consider the recent widely discussed evidence of cross-sample heterogeneity in decision biases provided by Krefeld-Schwalb et al. (2024), some of the best evidence yet for sensitivity to context in behavioral research. Closer examination of these results reveals that the default effect and framing replicate significantly in 10/11 populations and directionally in 11/11 populations, the less-is-better effect significantly and directionally in 11/11 populations, and the sunk cost effect significantly in 8/11 populations and directionally in 9/11 populations. There were precisely zero statistically significant reversals of any effect in any sample. Although researcher choices in experimental designs and statistical approaches can be hugely impactful (Landy et al., 2000; Silberzahn et al., 2018), there is not as much population variability in real findings as many scholars explicitly argue or implicitly assume.

More importantly, meta-analytic evidence reveals little cross-site heterogeneity in overall replication failures, sharply contradicting the context sensitivity defense (Olsson-Collentine, Wicherts, & van Assen, 2020). Effects that appear to be false positives under the criteria they fail to produce a significant directional estimate aggregating across sites (e.g., money or flag priming; Klein et al., 2014) are typically not characterized by significant replications at some sites, near-zero estimates at others, and significant reversals at others. The perspectivist thesis that most findings are massively moderated and thus likely to fully reverse across different populations (“the opposite of a great truth is also true”; McGuire, 1973, 1983) provides the intellectual backdrop for the hidden moderator rebuttal to failed replications. But although a

beautiful intellectual vision, perspectivism is not empirically supported by crowdsourced direct replications.

### *The expertise defence*

The accumulated empirical evidence also contradicts the claim that the modest overall replicability rate for published findings from top journals (Klein et al., 2014, 2018; Open Science Collaboration, 2015; Nosek et al., 2022) is attributable to replicator inexpertise (Bargh, 2012b; Baumeister, 2016; Schnall, 2014). Traditional indicators of scientific eminence such as publication records do not predict the empirical results replicators obtain (Bench et al., 2012; Landy et al., 2000). In addition, involving original authors as consultants or data collectors does not appreciably impact replication effect sizes (Klein et al., 2022; Schweinsberg et al., 2016), even when they resample the original population (Schweinsberg et al., 2016).

Some highly complex research paradigms (e.g., those involving confederates and hidden cameras) are much more difficult to scale than others, unquestionably limiting the scope of crowd science initiatives. Some laboratory measures, such as functional magnetic resonance imaging, require considerable prior training to deploy successfully. However, reasoning backward from a failed replication to the conclusion that “they must have done it wrong”, as is likely to occur in the case of Vaidis et al. (2024), is defensive, unscientific, and fallacious. If some findings are fragile and require expert hands, then more accomplished scientists by traditional metrics should be more likely to obtain them. But at least among the original effects re-examined thus far, it just isn't so.

The currently available evidence suggests that psychological findings are either fairly robust and generalizable across most research teams and participant populations (e.g., representativeness heuristic, defaults, framing, loss aversion; Klein et al., 2014; Krefeld-Schwalb et al., 2024) or they consistently are not (e.g., prime to behavior effects, ego depletion, effects of power poses on hormone levels, and so on; Cesario, Jonas, & Carney, 2017; Lodder et al., 2019; Klein et al., 2014; Verschuere et al., 2018; Vohs et al., 2021). The field seems to have produced one set of highly robust findings that hold across most contexts and another collection of dubious findings that do not emerge again and again when research is done under crowd conditions that put the expertise and context sensitivity arguments to systematic empirical tests.

#### *The operational failure defense*

As these earlier rebuttals face accumulating empirical counterevidence, the emergent defense against a systematic non-replication is now that of *operational failure* (Baumeister, Tice, & Bushman, 2023; Fiedler, McCaughey, & Prager, 2021). Perhaps the experimental manipulation did not successfully activate or affect the targeted mediating psychological state. If so, the replication may not have provided an informative test of the hypothesized causal relationship between the independent and dependent variable. As Baumeister et al. (2023) write, “Operational failures... do not constitute falsifications of the hypothesis, because they were unable to provide a test of it.”

Vaidis et al.’s (2024) addressed the operational failure concern by carefully measuring a key mediating state, specifically subjectively perceived choice. They find that the perceived voluntariness of writing a counter-attitudinal essay is greater in the high-choice condition, but

that this does not instigate attitude change as predicted by cognitive dissonance theory (Croyl & Cooper, 1983; Festinger & Carlsmith, 1959). Perhaps future replications should similarly capture mediating states, even when they were never assessed in the original study. Going further, even systematic crowdsourced replications could be collectively discounted by the scholarly community as uninformative if the manipulation does not significantly affect the mediator.

Manipulation checks and mediational measures are inherently valuable to include in both original studies and replications whenever feasible. However, the operational failure defense underestimates the severity of many skeptics' concerns about small sample classic studies. Indeed, there is one major form of meta-scientific skepticism regarding the original work that is supported, rather than undermined, by evidence of operational failure.

#### *Statistical skepticism vs. hypothesis skepticism*

One can distinguish between the *hypothesis skeptic*, who doubts the original theoretical claim (“It seems unlikely to me that perceived choice in engaging in a counter-attitudinal act causes attitude change”) and the *statistical skeptic* who dismisses implausibly large effect sizes from small underpowered studies as mainly noise rather than signal. Notably, the key meta-scientific papers that instigated the crisis of confidence in science focused principally on statistical and methodological concerns such as insufficient statistical power, effect size overestimation, researcher degrees of freedom, and publication bias — issues that generalize across research topics (Fanelli, 2010; Ioannidis, 2005; Simmons et al. 2011). Many meta-scientists and replicators, ourselves among them, approach the literature from the standpoint of a statistical skeptic rather than a hypothesis skeptic. We see limited informational value in an experimental



laboratory investigation with tiny numbers of participants per cell: the reported effects of condition on not only the dependent variable but also any process measures are at high risk of proving spurious.

There is therefore no need for a statistical skeptic to show that the ego depletion manipulation exhausted participants' mental resources, that fart spray made them feel disgusted, that recalling a time when they felt powerful made them feel powerful, or that the incidental presence of money activated thoughts about materialism. Given the statistical noise associated with the original designs, it is questionable whether these manipulations ever effectively induced their intended states or truly influenced scores on the dependent measures. Therefore, we would not expect either these mediating states or theorized downstream outcomes such as the ability to resist tempting treats, harsher moral judgments, more agentic behaviors, and greater cheating to prove robust. A replicator approaching the work from this stance need only repeat the original experimental manipulation and estimate the effect on the dependent variable using a large sample and pre-registered analyses. The inclusion of process measures could add value if the aim is to faithfully recreate the original experimental design in its entirety, but is not essential, especially if the original study itself featured no manipulation checks or mediational measures.

Another virtue of statistical skepticism, relative to hypothesis skepticism, is epistemological. From the perspective of traditional philosophy of science, it is extremely difficult to disprove scientific claims, especially in the social sciences (Kuhn, 1962; Lakatos, 1970; Lipton, 2008). In principle, an alternative operationalization of the independent or dependent variable could reveal support for the original theory. Thus, the hypothesis skeptic can deepen doubts, but never definitively falsify the original theoretical claim. Alternative experimental designs or variations

of the induced compliance paradigm might still demonstrate attitude changes under specific conditions, underscoring the provisional nature of hypothesis skepticism. In contrast, statistical skepticism focuses on the empirical robustness of findings rather than theoretical plausibility. There exist powerful tools capable of showing that a piece of experimental evidence does not provide robust positive support for the stated conclusions. If the original study reports implausibly large effects based on tiny samples (Schimmack, 2012), features p-values barely over the significance threshold (Simonsohn, Nelson, & Simmons, 2014; van Aert, Wicherts, & van Assen, 2016), and/or the effect systematically fails to emerge in numerous multisite direct replications, the narrow claim that the original work should be largely discounted in Bayesian terms is supported.

### *Conclusion*

The hypothesis skeptic suspects the original theory is false; the statistical skeptic suspects the original study captured mostly noise. Providing evidence against a theorized IV-DV link requires careful manipulation checks and measures of mediating states, as in Vaidis et al. (2024), and faces the potentially insurmountable epistemological and empirical challenge of proving that something never happens. In contrast, overwhelming the noisy estimates of unreliable original studies with the strong signals provided by superior multisite samples and more rigorous analyses is and should continue to be the primary goal of replication. This approach remains pivotal in advancing the reliability and validity of psychological research, revealing the clear signals of robust phenomena.

## **Funding/financial support statement**

This research was supported by an R&D grant from INSEAD to Eric Uhlmann.

## **Conflict of interests declaration**

Conflicts of interest: None.

## **References**

- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153-158.
- Baumeister, R.F., Tice, D.M., & Bushman, B.J. (2023). A review of multisite replication projects in social psychology: Is it viable to sustain any confidence in social psychology's knowledge base? *Perspectives on Psychological Science*, *18*(4), 912-935.
- Bargh, J.A. (2012a). Priming effects replicate just fine, thanks. *Psychology Today*. Retrieved April 11, 2017 at: <https://www.psychologytoday.com/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks>
- Bargh, J.A. (2012b). Nothing in their heads. *Psychology Today*. Withdrawn, but available at: <https://replicationindex.com/wp-content/uploads/2020/07/bargh-nothingintheirheads.pdf>
- Bench, S. W., Rivera, G. N., Schlegel, R. J., Hicks, J. A., & Lench, H. C. (2017). Does expertise matter in replication? An examination of the Reproducibility Project: Psychology. *Journal of Experimental Social Psychology*, *68*, 181-184.
- Cesario, J., Jonas, K.J., & Carney, D.R. (2017). CRSP special issue on power poses: What

was the point and what did we learn? *Comprehensive Results in Social Psychology*, 2, 1-5.

Croyle, R. T., & Cooper, J. (1983). Dissonance arousal: Physiological evidence. *Journal of Personality and Social Psychology*, 45(4), 782–791.

Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLOS ONE*, 5(4), e10068.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203-210.

Fiedler, K., McCaughey, L., & Prager, J. (2021). Quo vadis, methodology? The key role of manipulation checks for validity control and quality of science. *Perspectives on Psychological Science*, 16, 816–826.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science.” *Science*, 351, 1037.

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., Hasselman, F., Hicks, J. A., Hovermale, J. F., Hunt, S. J., Huntsinger, J. R., IJzerman, H., John, M., Joy-Gaba, J. A., Kappes, H. B., Krueger, L. E., Kurtz, J., Levitan, C. A., Mallett, R., Morris, W. L., Nelson, A. J., Nier, J. A., Packard, G., Pilati, R., Rutchick, A. M., Schmidt, K., Skorinko, J. L., Smith, R., Steiner, T. G., Storbeck, J., Van Swol, L. M., Thompson, D., van't Veer, A., Vaughn, L. A., Vranka, M., Wichman, A., Woodzicka, J. A., & Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, 45(3), 142–152.

- Klein, R.A., Vianello, M., Hasselman, F.,... Nosek, B.A. (2018). Many Labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490.
- Klein, R.A. et al. (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra*, 8(1), 35271.
- Krefeld-Schwalb, A., Sugerman, E. R., & Johnson, E. J. (2024). Exposing omitted moderators: Explaining why effect sizes differ in the social sciences. *Proceedings of the National Academy of Sciences*, 121(12), e2306281121.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. (1st Edn). Chicago, IL: University of Chicago Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave, Eds. *Criticism and the Growth of Knowledge* (pp. 91–195). Cambridge: Cambridge University Press.
- Landy, J. F., Jia, M., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johanneson, M., Pfeiffer, T., . . . & Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, 146(5), 451–479.
- Lipton, P. (2008). *Inference to the best explanation*. London: Routledge.
- Lodder, P., Ong, H. H., Grasman, R. P. P. P., & Wicherts, J. (2019). A comprehensive metaanalysis of money priming. *Journal of Experimental Psychology: General*, 148(4), 688-712.
- McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, 26(3), 446-456.
- McGuire, W.J. (1983). A contextualist theory of knowledge: Its implications for innovations

- and reform in psychological research. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 16, pp. 1-47). New York, NY: Academic Press.
- Nosek, B.A., Hardwicke, T.E., Moshontz, H., Allard, A., Corker, K.S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M.B. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748.
- Olsson-Collentine, A., Wicherts, J.M., & van Assen, M.A.L.M. (2020). Heterogeneity in direct replications in psychology and its association with effect size. *Psychological Bulletin*, 146(10), 922-940.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Schimmack, U. (2012). The Ironic Effect of Significant Results on the Credibility of Multiple-Study Articles. *Psychological Methods*, 17(4), 551-566.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90-100.
- Schnall, S. (2014). Social media and the crowd-sourcing of social psychology. Retrieved at: <https://www.psychol.cam.ac.uk/cece/blog>
- Schwarz, N., & Strack, F. (2014). Does merely going through the same moves make for a “direct” replication? Concepts, contexts, and operationalizations. *Social Psychology*, 45(4), 305-306.
- Schweinsberg, M., Madan, N., Vianello, M., Sommer, S. A., Jordan, J., Tierney, W., Awtrey, E., Zhu, L., Diermeier, D., Heinze, J., Srinivasan, M., Tannenbaum, D., Bivolaru, E., Dana, J., Davis-Stober, C. P., Du Plessis, C. Gronau, Q. F., Hafenbrack, A. C., Liao, E. Y., Ly, A., Marsman, M., Murase, T., Qureshi, I., Schaerer, M., Thornley, N., Tworek, C. M., Wagenmakers, E-J., Wong, L., Anderson, T., Bauman, C. W., Bedwell, W. L., Brescoll, V., Canavan, A., Chandler, J. J., Cheries, E.,

Cheryan, S., Cheung, F., Cimpian, A., Clark, M., Cordon, D., Cushman, F., Ditto, P. H., Donahue, T., Frick, S. E., Gamez-Djokic, M., Hofstein Grady, R., Graham, J., Gu, J., Hahn, A., Hanson, B. E., Hartwich, N. J., Hein, K., Inbar, Y., Jiang, L., Kellogg, T., Kennedy, D. M., Legate, N., Luoma, T. P., Maibeucher, H., Meindl, P., Miles, J., Mislin, A., Molden, D. C., Motyl, M., Newman, G., Ngo, H. H., Packham, H., Ramsay, P. S., Ray, J. L., Sackett, A. M., Sellier, A-L., Sokolova, T., Sowden, W., Storage, D., Sun, X., Van Bavel, J. J., Washburn, A. N., Wei, C., Wetter, E., Wilson, C., Darroux, S-C., & Uhlmann, E. L. (2016). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology, 66*, 55-67.

Silberzahn, R., Uhlmann, E. L., Martin, D., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T. J., Hederos, K., Heene, M., Hofelich Mohr, A. J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D., Lei, R., Lindsay, T. A., Liverani, S., Madan, C. R., Molden, D., Molleman, E., Morey, R. D., Mulder, L. B., Nijstad, B. A., Pope, N. G., Pope, B., Prenoveau, J. M., Rink, F., Robusto, E., Roderique, H., Sandberg, A., Schlüter, E., Schönbrodt, F. D., Sherman, M. F., Sommer, S., Sotak, K., Spain, S., Spörlein, C., Stafford, T., Stefanutti, L., Tauber, S., Ullrich, J., Vianello, M., Wagenmakers, E., Witkowiak, M., Yoon, S., & Nosek, B.A. (2018). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Advances in Methods and Practices in Psychological Science, 1*, 337–356.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False– positive psychology:

- Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76–80.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9, 666–681.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9, 59–71.
- Vaidis, D.C., Slegers, W.W.A., van Leeuwen, F., et al. (2024). A Multilab Replication of the Induced-Compliance Paradigm of Cognitive Dissonance. *Advances in Methods and Practices in Psychological Science*, 7(1).
- van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, 11(5), 713–729.
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, 113, 6454-6459.
- Van't Veer, A., & Giner-Sorolla, R. (2016). Pre-registration in social psychology: A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2-12.
- Verschuere, B., Meijer, E.H., Jim, A., Hoogesteyn, K., Orthey, R., et al. (2018). Registered Replication Report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, 1(3), 299–317.
- Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E.,



Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J.,... Albarracín, D. (2021). A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychological Science*, 32(10), 1566–1581.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J. & Kievit, R.A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632-638.