Psychon Bull Rev (2015) 22:1135–1141
DOI 10.3758/s13423-014-0757-4

BRIEF REPORT

Making sense of the noise: Replication difficulties of Correll's (2008) modulation of 1/f noise in a racial bias task

Christine Madurski · Etienne P. LeBel

Published online: 11 November 2014

Abstract Correll (Journal of Personality and Social Psychology, 94, 48–59, 2008; Study 2) found that instructions to use or avoid race information decreased the emission of 1/*f* noise in a weapon identification task (WIT). These results suggested that 1/*f* noise in racial bias tasks reflected an effortful deliberative process, providing new insights regarding the mechanisms underlying implicit racial biases. Given the potential theoretical and applied importance of understanding the psychological processes underlying implicit racial biases – and in light of the growing demand for independent direct replications of findings to ensure the cumulative nature of our science – we attempted to replicate Correll's finding in two high-powered studies. Despite considerable effort to closely duplicate all procedural and methodological details of the original study (i.e., same cover story, experimental manipulation, implicit measure task, original stimuli, task instructions, sampling frame, population, and statistical analyses), both replication attempts were unsuccessful in replicating the original finding challenging the theoretical account that 1/*f* noise in racial bias tasks reflects a deliberative process. However, the emission of 1/*f* noise did consistently emerge across samples in each of our conditions. Hence, future research is needed to clarify the psychological significance of 1/f noise in racial bias tasks.

Keywords 1/f noise · Implicit racial bias · Weapon identification task · Independent direct replication

With an increasingly multicultural and global society, the study of racial bias becomes ever more important. In this context, social psychologists have increasingly relied on implicit measures to assess individuals' racial attitudes – such as the Implicit Association Test (IAT) or the Weapon Identification Task (WIT) – which aim to overcome limitations of direct measures including socially desirable responding and introspective limits (Gawronski, LeBel, & Peters, 2007). Many of these implicit measures involve assessing individuals' reaction times (RTs) to a series of words or photos related to the attitude object (e.g., photos of African-American or Caucasian faces). For such tasks, RTs to different trial types are typically averaged across trials to minimize external influences on any one trial. Correll (2008) argued, however, that aggregating across trials ignores a great deal of information about the variation in trial-by-trial RTs and that considering such information from a 1/f noise perspective may shed new light about the psychological mechanisms underlying social psychological phenomena.

Correll (2008) investigated the potentially meaningful fluctuations in RTs across trials using an approach referred to as 1/*f* noise, which refers to non-random patterns of long-range correlations that manifest as waves in the fluctuations of RTs over time (Gilden, 2001; Gilden, Thornton, & Mallon, 1995; but see Wagenmakers, van der Maas, & Farrell, 2012). In recent years, 1/f noise – also known as flicker noise or pink noise – has been documented in a wide number of biological and physical systems including the fluctuations in tide heights, heartbeat, and firings of single neurons (Gilden, 2001; Press, 1978; for a review see Wijnants, 2014). From this perspective, the sequence of raw RTs can be represented as a complex waveform which can be decomposed into simpler component waves via a Fast Fourier transform (FFT). The log transformed frequency and power of each of these component waves can then be plotted; the slope between these two can then be estimated as power spectral density (PSD) slopes. If the variation in latencies is random then the PSD slope is not expected to differ from zero. However, PSD slopes that are negative, produced by lower frequency waves having more power than higher frequency waves, indicate 1/f noise. This suggests trial-to-trial variations in RTs are in fact non-random.

C. Madurski · E. P. LeBel (✉)
Department of Psychology, Dickson Hall, Montclair State University, Montclair, NJ 07043, USA
e-mail: etienne.lebel@gmail.com

Across two studies, Correll (2008) found that trial-by-trial variation in RTs revealed negative PSD slopes indicative of 1/*f* noise. In Study 1, Correll found that greater self-reported effort to avoid racial bias on a shooter task was correlated with less negative PSD slopes, and thus less 1/*f* noise. In Study 2, effort was experimentally manipulated. Participants instructed to use race or avoid race information while completing the WIT exhibited less negative PSD slopes than control participants (tested via a planned contrast whereby the average of the two experimental conditions had less negative PSD slopes than the control condition). The results suggested that 1/*f* noise in racial bias tasks reflects an effortful deliberative process, potentially providing new theoretical insights regarding our understanding of the nature of psychological processes underlying implicit racial biases (Fazio & Olson, 2003). Given the potential theoretical and applied societal importance of understanding the psychological processes underlying implicit racial biases – and in light of the growing demand for independent direct replications of findings to ensure the cumulative nature of our science (Koole & Lakens, 2012; Nosek, Spies, & Motyl, 2012), we decided to attempt to independently replicate Correll's Study 2 finding.[1]

Methods

In two large samples, we attempted to replicate Correll's (2008) Study 2 main finding using the exact same procedures, experimental manipulation, measures, stimuli, task instructions, sampling frame, and population. We contacted Correll to acquire any procedural and methodological details unreported in the published article and used large sample sizes to ensure high statistical power. Power analyses indicated that a sample size of 126 would be needed to achieve a power level of .80, based on the effect size of the critical contrast reported in the original study (*f*=.25, d=.59; power estimated using G-Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009). Given the availability of a large subject pool, however, we decided to aim for N=150 for both samples to provide even higher power levels. Furthermore, we also pre-registered our methods and planned statistical analyses prior to data collection to maximize transparency (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).[2]

For our first attempt, Correll (2008) provided all of the original stimuli for the WIT (White and Black male faces, tools, guns), the exact general instructions for the WIT, the exact instructions for each of the three conditions, and other methodological details not mentioned in the published article (i.e., trial order was randomized across participants, response key for "gun" and "tool" was on the right and left, respectively, and feedback for incorrect responses was presented for both practice and critical trials). We used the same sample type (laboratory sample) and sampling frame (undergraduate students participating for course credit).

For our second attempt, it was discovered that a smaller screen resolution and computer monitor was used compared to the original study, which made our stimuli appear about 23 % smaller than the stimuli in Correll's (2008) study. Therefore, in our second replication attempt, we increased the size of the stimuli by 32 % so that the stimuli appeared to our participants precisely the same size as they did to participants in Correll's original study.

For both replication attempts, however, there were two minor procedural differences. First, we used a standard keyboard to record responses rather than a response box as used by Correll because response boxes were not available in the laboratory rooms used. Second, a different beeping sound was used for incorrect responses because we used a different software than Correll.

Results

We analyzed the data following the exact same analytic approaches used by Correll (2008).[3] Indeed, we used the exact same SAS syntax included in the appendix of the original article to generate participant-specific PSD slopes via FFT from each participant's 200 trial-specific RTs. The main replication analysis involved a between-subjects ANOVA using a planned orthogonal contrast comparing the PSD slopes in the control condition to the average of the PSD slopes in the two experimental conditions (codes: control = -1, avoid race = +.5, use race = +.5).

As is shown in Fig. 1, we were not able to replicate Correll's (2008) Study 2 main finding in both of our samples.

Contrary to Correll's Study 2 finding, in both of our samples PSD slopes were not less negative in the use and avoid race conditions compared to the control condition (see Table 1).[4] Expectedly, however, mean PSD slopes in both samples were negative and statistically significantly

[1] We decided to attempt to replicate Correll's Study 2 finding because it provided the strongest test of the target hypothesis given it used an experimental manipulation, whereas Study 1 used a correlational design.

[2] Pre-registration involves specifying methodological and analytical plans in a frozen time-stamped document prior to data collection so that stringent confirmatory tests of the relevant hypotheses can be achieved (Wagenmakers et al., 2012b). Exact details of both replication attempts can be confirmed by cross-referencing the pre-registered replication protocols for replication attempt #1 and #2 available at https://osf.io/v3hfb/ and https://osf.io/czbzg/, respectively.

[3] In the spirit of open science practices, syntax files and de-identified data files for both of our replication attempts are available at https://osf.io/fejxb/ and https://osf.io/iraqy/.

[4] One participant in sample 2 had an extremely high error rate of 83 % and was excluded from all analyses. Including this participant yielded the same pattern of results, t(146)=-.42, p>.68, d=-.07.

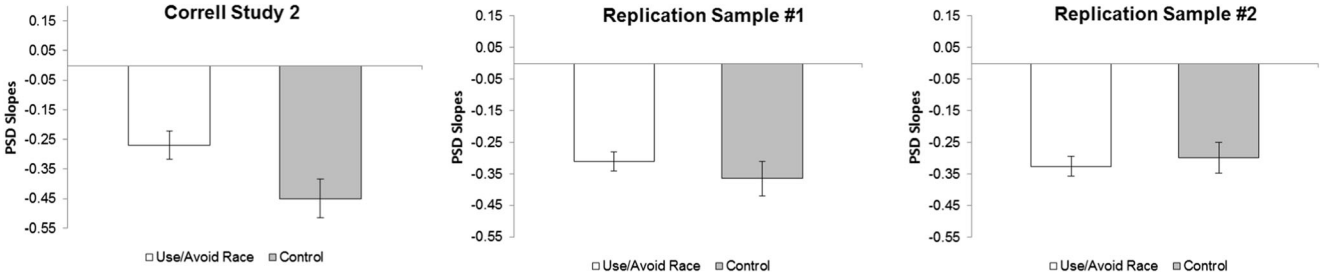

Fig. 1 Power spectral density (PSD) slopes across use/avoid race and control conditions in Correll's (2008, Study 2) original study and our two replication samples

different from zero in each of the effort instruction and control conditions (all *t*s<-6.10, all *p*s<.0001). Hence, our results did successfully replicate the standard 1/*f* noise pattern consistently found in past research (Torre, Balasubramaniam, Rheaume, Lemoine, & Zelaznik, 2011; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012) and as originally observed in Correll's (2008) control condition (see also Correll, 2011).[5]tgroup

We can gain additional clarity in interpreting our results via a Bayesian analysis, which quantifies the strength of evidence data provide for or against the null hypothesis relative to the alternative hypothesis (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Employing a Bayes Factor (BF) test for two-group designs using a non-informative Jeffrey-Zellner-Siow prior (Rouder, Speckman, Sun, Morey, & Iverson, 2009) revealed a BF of 9.92 for our combined sample (N=296) and a BF of .46 for Correll's (2008) Study 2 (N=71).[6] This indicates that our data provide about ten times more evidence for the null than the alternative hypothesis whereas Correll's data provide only about 2.2 times (inverse of .46) more evidence for the alternative than the null hypothesis. In other words, our replication results provide much more compelling evidence in favor of the null hypothesis than Correll's original evidence provides in favor of the effort decreases 1/f noise emission alternative hypothesis.

Discussion

Though 1/*f* noise did consistently emerge in each of our conditions across both samples – successfully replicating general 1/f noise patterns found in previous research (e.g., Torre et al., 2011; Wijnants et al., 2012) and specific conditions of Correll's prior work (i.e., Correll, 2008, Study 1 and control condition of Study 2; Correll, 2011) – we were unable to replicate Correll's Study 2 finding whereby instructions to use or avoid race information decreased the emission of 1/*f* noise. Our replication results are difficult to reconcile with Correll's original results for several reasons. Both of our samples were over twice as large as the one used by Correll, providing substantial statistical power to detect an effect comparable to the one reported by Correll (both samples having 86 % power, with the combined sample achieving 99 % power).[7] Of note, our combined analysis is in line with the *continuously cumulating meta-analytic* (CCMA) approach recently espoused by Braver, Thoemmes, and Rosenthal (2014). Additionally, our replication attempts were highly faithful to all procedural and methodological details of the original study (i.e., same cover story, experimental manipulation, implicit measure task, original stimuli, task instructions, sampling frame, population, and statistical analyses). Both replication attempts were also pre-registered, ruling out concerns regarding undisclosed flexibility in researcher degrees-of-freedom (LeBel et al., 2013; Simmons, 2011; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

Our results challenge Correll's (2008) theoretical account that 1/*f* noise in racial bias tasks reflects an effortful deliberative process, suggesting that more research is needed to clarify the psychological significance of the non-random 1/f noise pattern in racial bias tasks observed in our two samples and as originally observed by Correll (2008, 2011). Our results also speak to the continuing debate about the extent to which implicit racial bias measures (such as the WIT) are impervious to participants' intentional efforts to respond in ways that better mesh with their explicitly endorsed attitudes

[5] As suggested by a reviewer, observing 1/f noise in each of our conditions could have been interpreted as a successful replication of Correll's (2008) control condition if we had independent evidence that the instruction manipulation was unsuccessful in influencing exerted effort in our samples. We cannot ascertain this possibility, however, because an instruction manipulation check was not included, as was the case in the original study.

[6] These analyses were executed using Rouder et al.'s (2009) online calculator (http://pcl.missouri.edu/bf-two-sample) using the default scaling factor of r=1 and relevant *t*-values and *n*s (i.e., n1=47, n2=24, and *t*=2.35 for Correll's (2008) data and n1=198, n2=98, and *t*=.277 for our combined data).

[7] To further bolster our position, we also executed a safeguard-power-analysis (Perugini, Gallucci, & Costantini, 2014) on our combined sample to rule out concerns regarding imprecision in our power calculations due to the noisy effect size estimate in Correll's (2008) original study. This analysis revealed that we required an N=232 to reliably detect (80 % power) a lower bound effect size (d_s=.37) of Correll's observed effect size of d=.59 (R code for this analysis is available at https://osf.io/fejxb/ in "evaluating-replication-results.R").

Table 1 Critical contrasts of power spectral density (PSD) slopes between use/avoid race and control conditions in Correll's (2008, Study 2) original study and the current studies

Study	N	*t*	*p*	Effect size *d* +/- 95% C.I.	A priori power
Correll (2008, Study 2)	71	2.35	.02	d= .59 +/-.51	-
Current studies					
Sample #1	148	.891	.37	d= .16 +/-.34	86%
Sample #2	148	-.494	.62	d=-.09 +/-.34	86%
Combined	296	.277	.78	d= .03 +/-.25	99%

Note. A priori power refers to the probability of detecting an effect as large (or larger) than the original effect size of d=.59 as reported by Correll (2008, Study 2)

(Fazio & Olson, 2003). In this context, the general 1/f noise observed in our samples could be interpreted as being consistent with the theoretical position that implicit measures are not necessarily "process-pure" (Gawronski et al., 2007; Ranganath, Smith, & Nosek, 2008).

In interpreting our results, however, it is important to consider that our replication attempts did differ from Correll's (2008) original Study 2 in ways that may have contributed to the different results observed in our replication samples.

Different demographics

Nationality First, the demographics of our samples differed. Correll's (2008) sample consisted of American undergraduates, while our samples consisted of Canadian undergraduates. Given that African-American race-related biases have consistently been found in Canadian samples (e.g., Schuller, Kazoleas, & Kawakami, 2009), however, this demographic difference seems a priori an unlikely factor responsible for our different results. Furthermore, and more compellingly, behavioral evidence of racial bias (in terms of the number of stereotypically-congruent errors and RTs) was actually stronger in our samples than in Correll's original study. That is, participants instructed to use or avoid race exhibited higher levels of racial bias than participants in the control condition in both samples for RT bias and in one of our samples for error bias. On the other hand, neither of the bias indices were statistically significant across conditions in Correll's sample (see Table 2). These patterns of results suggest that Canadian participants had sufficient knowledge of the African-American stereotype, and hence nationality of sample is an unlikely explanation for our discrepant results.[8]tgroup

Ethnicity A closely related demographic variable that could have contributed to our discrepant results is a different ethnicity composition in our samples. This is unlikely, however, given that both of our samples and Correll's sample originated from large universities with a large proportion of international students.[9] Nonetheless, to rule this out we re-analyzed the target PSD slopes analysis including only White participants, but still failed to find a statistically significant difference across experimental and control conditions (Sample 1: $t(88)=1.55$, $p>.12$, d=.34; Sample 2: $t(85)=-.20$, $p>.83$, d=-.04; Combined sample: $t(176)=.96$, $p>.33$, d=.15).

Gender Another possibility is that our replication samples contained a different gender breakdown and this contributed to our discrepant results. Though possible, we contend this to be highly unlikely given that there is no known theoretical basis for expecting gender differences in racial biases. Furthermore, the gender composition in our samples was typical for psychology undergraduate students with a higher proportion of females than males (Sample 1 and 2 was composed of 65 % and 61 % females respectively; Correll did not report gender composition of his sample).

Non-compliance

Participant non-compliance could also have contributed to our different results. For instance, perhaps our participants did not follow instructions or responded carelessly during the WIT. However, allaying this concern is the fact that both of our studies revealed stronger behavioral evidence of heightened racial bias in the use/avoid race compared to the control condition than Correll (2008, Study 2). Nonetheless, to further rule out this concern, we specified conservative but reasonable non-compliance criteria (i.e., error rates greater than 20 % and mean RTs less than 200 ms) and re-analyzed the target PSD slopes analysis excluding participants meeting such criteria (N=11 and N=14 exclusions in Sample 1 and 2, respectively).

[8] That said, a reviewer raised a theoretically plausible possibility that our discrepant results may have been driven by the fact that Canadians may differ from Americans in their ability to control race bias given the dominant multicultural ideology of Canadian society. An empirical test of this interesting possibility awaits future research.

[9] Sample 1 and 2 ethnicity composition: 62 % Caucasian, 33 % Asians (incl. Indians), 2 % Blacks, and 3 % Other, and 60 % Caucasian, 30 % Asians (incl. Indians), 1.4 % Blacks, and 9.5 % Other, respectively. Correll (2008) did not report ethnicity composition and was not able to provide these upon request.

Table 2 Results of behavioral racial bias effects in Correll's (2008) sample and current replication samples

	Bias (# of errors)		Mean Diff. effect size [95 % C.I.]	Bias (in RTs, ms)		Mean Diff. Effect size [95 % C.I.]
	Use/Avoid race *M (SD)*	Control *M (SD)*		Use/Avoid race *M (SD)*	Control *M (SD)*	
Correll (2008, Study 2) (N=71)	1.30 (2.50)	0.63 (2.86)	d=.30 +/-.50 (*n.s.*)	9.67 (29.30)	-4.98 (38.43)	d=.43 +/-.51 (*n.s.*)
Current studies						
Sample # 1 (N=148)	1.78 (5.48)	0.06 (3.08)	d=.37 +/-.36 (*)	12.05 (67.57)	2.99 (46.99)	d=.34 +/-.35 (†)
Sample # 2 (N=148)	2.42 (5.79)	1.20 (4.62)	d=.22 +/-.35 (*n.s.*)	18.92 (61.74)	-4.80 (68.15)	d=.44 +/-.35 (*)
Combined (N=296)	2.10 (5.63)	0.60 (3.96)	d=.28 +/-.25 (*)	15.45 (64.68)	-0.98 (58.58)	d=.39 +/-.25 (*)

Note. * = $p<.05$, † = $p<.055$, *n.s.* = not statistically significant, RT = Reaction time. Mean difference effect size (with +/- 95 % confidence interval. and statistical significance) reflect mean differences in racial bias scores in use/avoid race condition compared to control condition. Following Correll (2008), bias in terms of # of errors was calculated as: # of errors on Black-tool trials minus # of errors on White-tool trials plus # of errors on White-gun trials minus # of errors on Black-gun trials. Bias in terms of RTs was calculated analogously as: RT on Black-tool trials minus RT on White-tool trials plus RT on White-gun trials minus RT on Black-gun trials. For both racial bias indices, larger values are assumed to reflect higher levels of racial bias. For Bias in RTs (and again following Correll), effect size and statistical significance were calculated on log-transformed RTs, with raw RTs reported for ease of interpretation.

PSD slopes across experimental and control conditions were still not statistically significant excluding these participants, further bolstering our case that non-compliance cannot explain our discrepant results (Sample 1: $t(134)=1.08$, $p>.28$, d=.20; Sample 2: $t(131)=-.59$, $p>.56$, d=−.11; Combined sample: $t(268)=.36$, $p>.72$, d=.05).

Another possibility is that because our participants were run in groups of two to five (rather than individually as in Correll, 2008), they could have been distracted by the presence of the other participants. We believe this possibility to be unlikely given that great care was taken to minimize distractions by seating participants in separate partitioned cubicles. Participants were also wearing headphones. Additionally, the experimenter seated participants in the cubicles furthest from the door first, to avoid the possibility that tardy participants distract participants already completing the study.

Poor psychometric properties

Yet another possibility is that the psychometric properties of the WIT in our replication samples were somehow different from Correll's (2008) sample or substandard. However, reliability estimates for WIT scores were $\alpha=.53$ and $\alpha=.54$ in our first and second samples, respectively, which are reasonable for implicit measures (LeBel & Paunonen, 2011) and substantially higher than in Correll's sample ($\alpha=-.21$).[10] Hence, poor psychometric properties cannot account for the discrepant results observed in our replication attempts.

[10] Following standard procedures for implicit measures (LeBel & Paunonen, 2011), reliability estimates were estimated using a split-half approach whereby separate WIT scores were calculated for even- and odd-numbered trials and a Cronbach's alpha calculated on both of these halves.

Hardware differences

Minor differences in the hardware used in our replication studies could also have contributed to the different results observed. For instance, we used slightly different computer monitor sizes, which could have affected the actual size that the stimuli appeared to our participants. Indeed, as mentioned, it was discovered after our first replication attempt that the stimuli appeared approximately 23 % smaller to our participants given that we used larger computer monitors with a higher screen resolution than Correll (2008). For our second replication attempt, the size of the stimuli was increased by 32 % so that they appeared the same physical size to our participants as in Correll's study. However, given that our second replication attempt also failed to replicate Correll's original finding, it is unlikely that this hardware difference can explain our discrepant results.

Another minor hardware difference was that we used a keyboard whereas Correll (2008) used a response box. However, given that standard keyboards are typically accurate to about +/- 7.5 ms (Segalowitz & Graves, 1990), this hardware difference is also unlikely to have had any significant effects on the obtained results.

A final minor hardware difference was the beeping sound used for incorrect responses. This difference is unlikely to account for our discrepant results, however, given that the beeping sound was a standard beeping sound approved by Correll prior to data collection (it was necessary to use a different beeping sound because we used a different software than Correll).

In summary, despite considerable effort to duplicate all of the procedural and methodological details of the original study, two high-powered pre-registered replication attempts were unsuccessful in corroborating Correll's (2008, Study 2) finding whereby instructions to use or avoid race information

reduced the emission of 1/f noise.[11] That being said, our negative results do not necessarily rule out the possibility that effort instructions could influence the emission of 1/*f* noise in a different context, under different conditions (e.g., many more trials per subject), or under a different set of operationalizations, each of which could be identified in future research. For instance, alternative scaling methods could be used to examine 1/f noise such as detrended fluctuation analysis (DFA) or standardized dispersion analysis (SDA), which have been argued to yield more robust results with relatively short time-series data (Hasselman, 2013). Of more theoretical importance, however, we did consistently observe general patterns of 1/*f* noise in each of our conditions across both samples – successfully replicating general 1/f noise results observed in past research (Torre et al., 2011; Wijnants, 2014; Wijnants et al., 2012) and as originally observed in specific conditions of Correll's prior work (i.e., Correll, 2008, Study 1 and control condition of Study 2; Correll, 2011). Consequently, it is important to emphasize that though our results challenge Correll's (2008) theoretical account that 1/*f* noise in racial bias tasks reflects an effortful deliberative process, our results corroborate the fact that 1/f noise does indeed emerge in implicit racial bias tasks. Hence, clarifying the psychological significance of such non-random 1/f noise pattern represents an intriguing puzzle for future research to clarify.

Acknowledgments We would like to thank Joshua Correll for his cooperation in providing study materials and other methodological details and Amanda Abado and Monica Bochus for help with data collection. We also thank Hal Pashler, Mark Brandt, Marco Perugini, Joe Cesario, Lorne Campbell, Fred Hasselman, and E-J Wagenmakers for valuable feedback on an earlier version of this manuscript. This research was partially supported by a Social Science and Humanities Research Council (SSHRC) post-doctoral fellowship to EPL.

References

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9,* 333–342.

Correll, J. (2008). 1/*f* noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology, 94,* 48–59.

Correll, J. (2011). Order from chaos? 1/*f* noise predicts performance on reaction time measures. *Journal of Experimental Social Psychology, 47,* 830–835.

[11] Our replication results can also be considered "conclusive failures" to replicate Correll's (2008, Study 2) original finding according to Simonsohn's (2013) small-telescope approach. From this perspective, our combined-sample d=.03 was statistically significantly smaller than an objectively detectable small effect in the original study ($d_{33\,\%}$=.26). (R code for this analysis is available at https://osf.io/fejxb/ in "evaluating-replication-results.R").

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297–327.

Hasselman, F. (2013). When the blind curve is finite: dimension estimation and model inference based on empirical waveforms. *Frontiers in Physiology*, *4*. DOI: 10.3389/fphys.2013.00075

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Association for Psychological Science, 2,* 181193.

Gilden, D. L. (2001). Cognitive emissions of 1/*f* noise. *Psychological Review, 108,* 35–56.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/*f* noise in human cognition. *Science, 267,* 1837–1839.

Koole, S. L., & Lakens, D. (2012). Rewarding replications a sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7,* 608–614.

LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots support for reforming reporting standards in psychology. *Perspectives on. Psychological Science, 8,* 424–432.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin, 37,* 570583.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7,* 615–631.

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. Perspectives on Psychological Science, 9(3), 319-332.

Press, W. H. (1978). Flicker noises in astronomy and elsewhere. *Comments in Astrophysics, 7,* 103–119.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44,* 386–396.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237.

Schuller, R. A., Kazoleas, V., & Kawakami, K. (2009). The impact of prejudice screening procedures on racial bias in the courtroom. *Law and Human Behavior, 33,* 320–328.

Segalowitz, S. J., & Graves, R. E. (1990). Suitability of the IBM, XT, AT, and PS2 keyboard, mouse and game port as response devices in reaction time paradigms. *Behavior Research Methods, Instruments, & Computers, 22,* 283–289.

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.

Simonsohn, U. (2013). Small telescopes: Detectability and the evaluation of replication results (December 10, 2013). Available at SSRN: http://ssrn.com/abstract=2259879

Torre, K., Balasubramaniam, R., Rheaume, N., Lemoine, L., & Zelaznik, H. N. (2011). Long-range correlation properties in motor timing are individual and task specific. *Psychonomic Bulletin & Review, 18,* 339–346.

Wagenmakers, E. J., van der Maas, H. L., & Farrell, S. (2012). Abstract concepts require concrete models: Why cognitive scientists have not yet embraced nonlinearly coupled, dynamical, self-organized critical, synergistic, scale-free, exquisitely context-sensitive, interaction-dominant, multifractal, interdependent brain-body-niche systems. *Topics in Cognitive Science, 4,* 87–93.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100,* 426–432.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012b). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7,* 627–633.

Wijnants, M. L. (2014). A review of theoretical perspectives in cognitive science on the presence of 1/f scaling in coordinated physiological and cognitive processes. *Journal of Nonlinear Dynamics*. doi:10.1155/2014/962043

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia, 62,* 100–119.