



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Evidence against subliminal anchoring: Two close, highly powered, preregistered, and failed replication attempts

Lukas Röseler^{a,b}, Astrid Schütz^{a,*}, Pia A. Blank^a, Marieluisa Dück^a, Sabine Fels^a, Jana Kupfer^a, Linda Scheelje^a, Christian Seida^a^a University of Bamberg, Germany^b Harz University of Applied Sciences, Germany

A B S T R A C T

When people consider an arbitrary number prior to generating a numeric estimate, their estimate is typically biased toward that number. This phenomenon is called anchoring and has been described as one of the most robust phenomena in judgment and decision making. However, the literature on anchoring has been plagued by numerous blatant contradictions, one of which is characterized by opposing statements regarding the question of when subliminal anchors work (i.e., under time pressure vs. when people take their time). We address this inconsistency by replicating two studies using high-powered direct replications and preregistration based on the “replication recipe” (Brandt et al., 2014). We could not find any evidence of subliminal anchoring in either of the two replications.

When numeric estimates are biased toward a previously considered numeric value, an anchoring effect has occurred. This phenomenon has received remarkable attention ever since Tversky and Kahneman (1974) described it. Numerous theories have been proposed (e.g., Frederick & Mochon, 2012; Mussweiler & Strack, 1999a), many moderators have been investigated (for an overview, see Furnham & Boo, 2011), and classical effects have been successfully replicated with very large effect sizes (e.g., Klein, Ratliff, Vianello, Adams and Bahník, 2014). The limits and mechanisms of anchoring are still being investigated, though (Frech, Loschelder, & Friese, 2020; Lewis, Gaertig, & Simmons, 2019). One unresolved issue concerns the question whether an individual needs to consider an anchor to become biased or whether mere perception (without deliberate processing) of anchors is sufficient (e.g., Mussweiler & English, 2005).

1. Subliminal anchoring

Subliminal anchoring is a special case of anchoring in which an anchor is not processed but presented subliminally. The anchor is not explicitly mentioned and subsequently considered by the respondent as in other paradigms that typically feature the comparative question “Is it more or less than X?” To our knowledge, there are only two published studies that have investigated *subliminal anchoring*. Both studies found subliminal anchoring effects. However, the conditions that have been proposed to be necessary for subliminal anchoring are contradictory. Mussweiler and English (2005) provided the first published evidence of subliminal anchoring. As part of their cover story, they asked

participants to concentrate and *take their time* in arriving at an estimate. During this period of concentration, the anchor was presented. By contrast, Reitsma-van Rooijen, Daamen, and L. (2006) reported that subliminal anchoring works only when participants are under time pressure, which is the opposite of what Mussweiler and English suggested.

Neither paper provided a thorough theoretical account for why (no) time pressure should be essential or how subliminal anchors actually work because both were rather exploratory approaches. The selective accessibility model of anchoring (e.g., Mussweiler & English, 2005; Mussweiler & Strack, 1999b) can be applied to subliminal anchoring because it uses priming as one of its mechanisms. However, given the recent failure to replicate one of its key findings (Harris et al., 2019) and its contradictory predictions in some domains, it does not seem to provide a clear theoretical account of subliminal anchoring. For example, ego depletion is predicted either to increase (Banker, Ainsworth, Baumeister, Ariely, & Vohs, 2017) or to decrease (Francis, Milyavskaya, Lin, & Inzlicht, 2018) the strength of anchoring on the basis of the selective accessibility model, but there was no effect at all in a registered report (Röseler, Schütz, Baumeister, & Starker, 2020). Based on the arguments provided for the positive or negative effect of ego depletion on the strength of anchoring, stronger effects (less time leads to stronger susceptibility to situational cues; e.g., Banker et al., 2017) and weaker effects (more time leads to more consideration and stronger priming; e.g., Francis et al., 2018) can be predicted for the presence of time pressure versus the absence of time pressure in subliminal anchoring. With the present paper, we are attempting to clarify the

* Corresponding author.

E-mail address: astrid.schuetz@uni-bamberg.de (A. Schütz).

contradiction with respect to time pressure between the two contradictory studies so that new models can account for the role of time pressure.

If one of these studies cannot be replicated, the contradiction will be solved. Thus, we conducted high-powered preregistered replications of the two subliminal anchoring studies. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012).

2. Study 1: replication of Mussweiler and English (2005), Study 2

Mussweiler and English (2005) demonstrated anchoring effects when participants concentrated on a computer screen for 1 min and anchors were subliminally presented 10 times for 33 ms each time and masked with symbols (e.g., $\$?#\beta\#$). High and low anchors influenced the estimate of the mean temperature in Germany (Study 1) and the price of a middle-class car (Study 2). Apart from anchors and the target question, the only other difference between Mussweiler and English's Studies 1 and 2 was the prime detection task in Study 2, in which participants tried to recognize the subliminally presented anchor. Study 2 showed that participants could not recognize subliminal anchors even when participants knew the anchors existed and wanted to recognize them. We conducted a high-powered, preregistered, close replication of Study 2 in which participants estimated the average price of a middle-class car.

2.1. Method

As in the original study, in the current study, participants were presented with a question (the average price of a middle-class car) and were then asked to focus on flickering letters (among which the anchor was presented) and subsequently to provide an estimate regarding the target question. We determined the anchors based on a pretest. The high anchor was 40,000 € and the low anchor was 10,000 € (anchors in the original study were 30,000 € and 10,000 €; the data from the pretest is available online, <https://osf.io/8bqmy/>). Afterwards, participants completed a funneled debriefing and a prime detection task. We used the funneled debriefing method to test for participants' awareness of the primes, which included seven awareness check questions and finally revealed the purpose of the experiment (Mussweiler & English, 2005, p. 136).¹ In the prime detection task, participants were told about the subliminal anchors and were asked to guess the numbers presented in 10 trials (either 10,000 or 40,000). After each trial, participants indicated which number they believed to have seen by pressing a button on the keyboard. The result was used to test whether their performance was better than chance and whether the presentation was actually subliminal. The presented numbers were randomized such that the high anchor (40,000) was presented five times, and the low anchor (10,000) was presented five times. For an overview of the procedure of the experiment and the anchoring task see Fig. 1. In a personal correspondence, the first author of the original study revealed that the original materials were no longer available (the original study was conducted more than 14 years ago). We thus designed our study to resemble the original study as closely as possible using the details described in the original publication. We preregistered the replication before data collection using the replication recipe (Brandt et al., 2014).

¹ The seven items that we took from Mussweiler and English (2005) were "(1) did you notice anything special in this study? (2) what do you think this study was about? (3) did you notice anything special with the fixation string? (4) did you notice that presentation of this letter string was interrupted? (5) do you have any idea of what the interruptions consisted? (6) in fact, the fixation letter string was interrupted by the very brief presentation of numbers. Were you able to detect these numbers? (7) please write down the numbers you detected. (p. 136)"

2.1.1. Planned sample size

To facilitate a power analysis, we pooled the effect sizes of the original Studies 1 and 2 (which differed only with respect to the target question) and found a mean effect of $d = 0.69$ ($N = 76$, 95% CI = [0.23, 1.16], $N_{min95\%} = 94$) of subliminal anchoring. The present study was conducted as part of a research class and thus the target sample size was also chosen for practical reasons. We decided to collect data from $N = 160$ participants (more than two times the sample size of Mussweiler and English's studies 1 and 2 or 3.8 times the sample size of Mussweiler and English's study 2), which would yield $1 - \beta > 0.99$ for $d = 0.69$ (original effect size) and $1 - \beta > 0.80$ for $d = 0.40$ (mean effect size for a variety of anchoring effects).

2.1.2. Deviations from the original study

Our study deviated from the original study in several ways. (a) We could not conduct the study in a lab, but we instructed participants to do the experiment in a calm working environment. Participants were recruited among psychology students by announcing and advertising the study over students' networks. Some students completed the study in the University computer lab, some did the study at home. Participants were unaware of the goals of the study. After completion of the study, participants who participated at home were informed how to determine their monitors refresh rate and asked to report it - as this was essential to conduct the study as planned. All participants reported a refresh rate of 60 Hz. In the computer laboratory refresh rates were likewise 60 Hz.

(b) To achieve our target sample size, we recruited not only students but also nonstudents. (c) We did not reward participants as had been done in the original study by giving participants 6 € (Study 1) or an ice cream cone (Study 2) after the study. However, those participants had probably not been told about the ice cream beforehand, and the provision of the cone was not linked to performance. Thus, not providing a reward can be viewed as a minor deviation in terms of replication closeness. (d) As in the original study, we conducted a small pretest to determine the high and low anchor values. The mean estimate of the price of a middle class car was $M = 23,026.32$ € ($SD = 6441$, $N = 19$), which is why we chose 10,000 and 40,000 as anchors. (e) We decided to document whether the participants were students to see if there was a difference between students (i.e., the participants in the original study) and nonstudents.

2.1.3. Exclusion criteria

In Mussweiler and English's (2005) Studies 1 and 2, the published exclusion criteria were ambiguous. They gave two reasons for the exclusions. (a) A participant who recognized all 10 subliminal anchors in the prime detection task was removed from the sample because it could not be guaranteed that the subliminal anchor presentation had worked. (b) Two participants who "indicated their suspicion about the flickering of the letter string" (p. 136) were also removed. In our preregistration, we specified this exclusion criterion as an answer of "yes" to Funnel Debriefing Question 4, that is, "Did you notice the flickering?" While collecting our data, however, we noticed that this exclusion criterion resulted in one third of the sample being excluded (in the original study, it resulted in only 2 of 44 = 4.5%). Furthermore, we noticed that excluding somebody because he or she noticed that there was flickering was problematic because participants were explicitly told about the flickering in the original instructions: "To remind them to keep focus, the letter string would flicker regularly" (Mussweiler & English, 2005, p. 135). We thus deviated from our preregistration and instead excluded participants who answered yes to Funnel Debriefing Question 6, which was, "There were numbers interspersed with the letters. Could you tell what the numbers were?" This change led to far fewer exclusions (see results) but did not change the results of the hypothesis test and manipulation check.

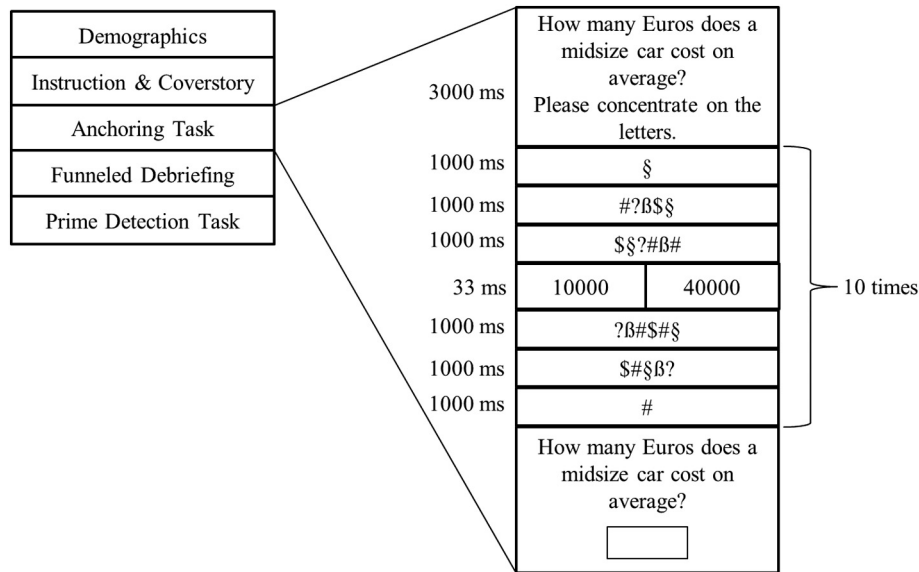


Fig. 1. Procedure of our Study 1 and the anchoring task.

2.1.4. Definition of a successful replication

We defined the replication attempt as successful if both of the following criteria were met: (a) The effect of the subliminal anchors was significantly larger than 0, and (b) the effect was not significantly smaller than the original effect [for the sake of precision, we pooled [Mussweiler & English's, 2005](#) effect sizes from Studies 1 and 2, $d = 0.69$].

2.2. Results

2.2.1. Achieved sample size

A total of 183 participants completed the experiment. Six participants were excluded because they reported all 10 subliminal anchors correctly; 64 participants would have been excluded because they noticed the flickering on the screen (for a discussion, see the Method section); 12 participants were excluded because they said they had recognized the subliminal anchors. Of the remaining 165 participants, 80 had been presented with the high and 85 with the low anchor; 93 were women, 67 men, and 3 diverse; 136 were students. The mean age was $M = 24.12$ years ($SD = 6.95$ years). The achieved power for $d = 0.69$ (pooled original effect size) was thus $1 - \beta > 0.999$.

2.2.2. Manipulation check and hypothesis test

In the prime detection task, participants' performance was better than chance, that is, they detected significantly more than 5 of 10 primes ($M = 5.73$, $SD = 1.66$), $t(164) = 5.63$, $p < .001$, $d = 0.440$. This indicates that anchors were not presented subliminally to all participants. However, the estimates in the high anchor condition ($M = 19,374.03$, $SD = 24,112.77$, $N = 80$) were not significantly higher than the estimates in the low anchor condition ($M = 24,384.96$, $SD = 41,229.61$, $N = 85$), $t(135.03) = -0.96$, $p = .830$ (one-tailed), $d = -0.148$. This effect was significantly smaller than the pooled original effect ($d = 0.69$). The combined effect (i.e., the pooled original effect and the replication effect) was $d = 0.12$ ($N = 241$) and not significant, either, $p = .37$ (two-tailed).

2.3. Exploratory analyses

All further analyses were not preregistered and were conducted for exploratory purposes only. Beware of the high probability of false positives and false negatives given the high number of exploratory analyses and the relatively small sample size.

2.3.1. Excluding outliers

There were five outliers in both the high and the low anchor group ($Max = 300,000$, $Z_{max} = 8.19$). In an analysis that was not preregistered, we applied an exclusion criterion that is common in anchoring research, that is, we excluded values that were three standard deviations above or below the mean estimate (e.g., [Chaxel, 2014](#), $p. 47$). The estimates in the high anchor condition ($M = 20,763.89$, $SD = 11,136.08$, $N = 65$) were not significantly larger than the estimates in the low anchor condition ($M = 21,100.00$, $SD = 9949.01$, $N = 71$), $t(128.82) = -0.18$, $p = .573$ (one-tailed), $d = -0.032$.

2.3.2. Susceptibility to anchoring and prime detection task

[Mussweiler and English \(2005\)](#) tested for a correlation between the susceptibility to the subliminal anchor (i.e., absolute difference between the estimate and anchor) and the number of correctly recognized primes in the prime detection task. As in the original study, we did not find a correlation, $r(134) = 0.049$, $p = .572$ (two-tailed). This result was not sensitive to including or excluding the outliers.

2.3.3. Strength of anchoring effects and detection of the flickering

A large proportion of participants detected the flickering (55 of 165 participants). We tested for whether anchoring effects were stronger or weaker for these participants than for those who did not detect the flickering. The interaction between the anchor (low vs. high) and flickering detection (yes vs. no) was not significant, $F(1,132) = 0.76$, $p = .386$. This result was not sensitive to keeping or removing the outliers.

2.3.4. Participant age

By including nonstudents, our sample's mean age was most likely higher than the mean age from the original study (but [Mussweiler & English, 2005](#), reported no demographic data). It is possible that the optimal calibration of the subliminal anchor's presentation time further depends on participants' age. We thus tested for whether age was correlated with the number of correctly recognized subliminal anchors in the prime detection task, but we found a very small and nonsignificant correlation, $r(161) = -0.152$, $p = .053$ (two-tailed).

2.4. Discussion

Our high-powered, preregistered, close replication of [Mussweiler and English \(2005\)](#) may have been miscalibrated because the anchors were not necessarily presented subliminally. It is, however, surprising

that participants recognized the anchors that were presented for 33 ms because the exposure times for subliminal primes are usually up to 500 ms (Elgendi et al., 2018, p. 23). Moreover, there was no effect, and mean estimates in the high anchor condition were even lower than in the low anchor condition. Due to ambiguous reports in the original study, we changed one preregistered exclusion criterion but found that whether or not the participants recognized the flickering had no impact on the strength of the anchoring.

To sum up, due to the extremely high power, the anchoring effect should have occurred if it indeed existed. We interpret the fact that there was no sign of anchoring—even after we excluded outliers and controlled for whether participants saw the flickering—as inconsistent with the claim that subliminal anchoring exists.

Two claims in favor of subliminal anchoring can be made at this point. (1) Subliminal anchoring might only occur under time pressure. (2) Due to the partly supraliminal presentation of the anchors and the partly uncontrolled test environment, Study 1 may not provide a strict test of the subliminal anchoring hypothesis: If some people saw the anchor, their awareness might have led them to over correct their estimate because they felt manipulated (e.g., Wilson, Houston, Etling, & Brekke, 1996, p. 400). Study 2 addresses both problems by introducing a condition with time pressure and reducing the presentation time of the anchors to 17 ms instead of 33 ms.

3. Study 2: replication of Reitsma-van Rooijen et al. (2006)

Our results are not the first to call into question Mussweiler and English's (2005) claim about the existence of subliminal anchors. One year after Mussweiler and English's study came out, Reitsma-van Rooijen et al. (2006) published results of an experiment on subliminal anchoring and reported that subliminal anchoring occurred only under time pressure. In their experiment, which was slightly different from the one by Mussweiler and English, participants were first exposed to the subliminal anchors (10 vs. 90) for only 17 ms (instead of 33 ms in the study by Mussweiler & English, 2005) and then asked to provide an estimate of the probability of the recurrence of a pestilence of the lungs in India within a year. Participants were asked/not asked to provide their answer quickly. Anchoring occurred only in the condition in which participants had to give their answer quickly. Our null findings from the replication of Mussweiler and English's Study 2, in which participants were asked to take their time, are consistent with the findings of Reitsma-van Rooijen and Daamen. Together this suggests that subliminal anchoring occurs only under time pressure. We thus conducted a high-powered, preregistered, close replication of the study by Reitsma-van Rooijen and Daamen.

3.1. Method

In the present study, just as in the study by Reitsma-van Rooijen et al. (2006), participants were greeted and then asked to complete demographic questions (gender, age, and whether they were students). Then they were told that this study was about how people make estimates and that “we knew from experience that participants were able to give the correct answer, when they followed their first impression” (Reitsma-van Rooijen et al., 2006, p. 383). They were introduced to the letter task, which required them to say whether a random letter combination consisted of more lower or more upper case letters. Anchors were subliminally presented for 17 ms during the letter task (high anchor: 90, low anchor: 10). After the letter task, they were asked to provide an estimate of the probability of the recurrence of a pestilence of the lungs in India within a year. See Fig. 2 for an overview of the procedure. One difference was that participants who had completed the study in our lab were given candy. Furthermore, we preregistered the replication before data collection using the replication recipe (Brandt et al., 2014).

3.1.1. Letter task and presentation of anchors

In the original study, there was an exercise to familiarize participants

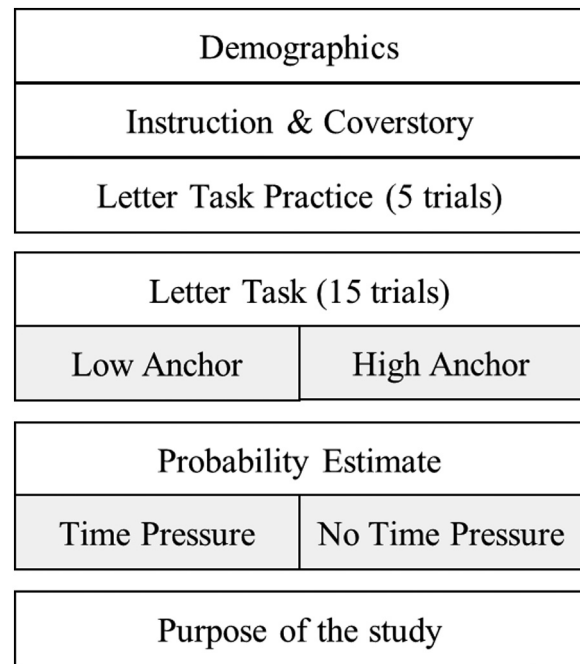


Fig. 2. Procedure of Study 2.

with the letter task. As this exercise was not described and we were unable to obtain information from the authors, we created a new one using the description in the publication: Participants completed five trials of the very same task that they would encounter later, except that there were no subliminal anchors. In the letter task, (a) two crosses (“XX”) were presented for 500 ms in the center of the screen, (b) then the anchor (low anchor: 10; high anchor: 90) was presented for 17 ms (this was omitted in the exercise), (c) the two crosses were presented for 17 ms, (d) the letter combination was presented for 1500 ms, and then (e) the question appeared “Were there more upper or lower case letters?” until participants gave their answer by pressing either A (upper) or L (lower) on the keyboard. After 15 trials, participants saw the message “This was the last letter combination; please make the probability estimate [please give your answer quickly]” for 5000 ms. Afterwards, participants typed their estimate. The request for a quick answer was omitted from the no-time-pressure condition.

3.1.2. Letter task stimuli

The original study used Chicago font, plain, 14 pt., 4.5 mm high from a 0.5 m viewing distance. Because Chicago font is no longer used on contemporary computers, we used a similar sans serif font that is common across most computers, namely, Tahoma. We made sure that letters were 4.5 mm high and were viewed from a 0.5 m viewing distance in the lab. We also excluded upper case and lower case letters that were likely to be confused, such as c/C, k/K, and v/V. A list of all the letters we used can be found online (see links in the open practices section). In the original study, there were 15 letter combinations, and “there were as many capital letters as lower case letters added over the letter combinations” (Reitsma-van Rooijen et al., 2006, p. 383). Due to a lack of further information, we decided that all letter combinations would have a 4:2 or 2:4 ratio of upper:lower case letters. Using pseudo-random numbers, we generated 8 combinations with a 4:2 ratio and 8 combinations with a 2:4 ratio. Each participant saw 15 of these 16 combinations. For the exercise, we created 2 times 3 combinations using the same ratios.

3.1.3. Planned sample size

The original effect sizes were $d = -0.284$ for the no-time-pressure condition and $d = 0.867$ for the time-pressure condition. A total of 62 participants (15 or 16 participants per cell) completed the original study. As the authors of the original study found that there was no effect in the no-

time-pressure condition, we predicted an interaction effect size of $f = (d1-d2)/4 = (0.867-0)/4 = 0.21675$. Because achieving 95% power ($N_{\min} = 279$) was not possible for practical reasons, we aimed for 80% power and thus a target sample size of $N = 170$, which was still more than two times the original sample size.

3.1.4. Deviations from the original study

We deviated from the original study in several ways. (a) We did not present the question for 5000 ms but for as long as participants needed to type in their estimate and click on the “next” button ($Mdn_{response\ time} = 18.36\ s$). (b) We designed a new exercise and new stimuli for the letter task. (c) We asked for demographic data (i.e., gender, age, and student status). (d) We applied preregistered exclusion criteria. (e) Finally, due to the SARS-CoV-2 pandemic, we collected some of our data online rather than in the lab.

3.1.5. Exclusion criteria

As percentages have to be between 0 and 100, all participants who estimated a number below 0 or above 100 were excluded. Second, if participants did not concentrate on the letter combinations, they missed the subliminal anchor that was presented before the letters. Thus, not being able to report the correct letter combination was an indicator of careless responding. This is why we excluded participants if they provided wrong answers in 12 or more of the 15 letter task trials. The five trials from the exercise were not considered.

3.1.6. Definition of successful replication

We considered the replication attempt successful when (a) the interaction effect size was significantly larger than $f = 0$ and (b) the interaction effect was not significantly smaller than $f = 0.22$ (i.e., the original study's interaction effect size).

3.2. Results

3.2.1. Achieved sample size

Due to the Coronavirus pandemic, which changed everyday life around the world, we had to stop recruitment in our laboratory and continue the survey online (thereby deviating from the preregistration). A total of 111 participants were administered the original version in our labs, whereas 59 participants took the online version; 175 participants completed the experiment (see links in the open practices section). No participant was excluded because of estimated probabilities larger than 100%. No participant was excluded due to too few correct responses in the letter task. Three participants were excluded because they recognized that the experiment was about anchoring or subliminal perception. The remaining sample was reduced to the target sample size of $N = 170$, and 80% power was thus achieved. Participants consisted of 121 women, 48 men, and 1 diverse; 114 were students. Their mean age was $M = 31.97$ years ($SD = 17.19$ years).

3.2.2. Manipulation check

To test whether asking participants to give the answer quickly actually made participants provide the answer quickly, we analyzed whether estimates in the time-pressure condition were faster than in the no-time-pressure condition. We therefore log-transformed the estimates as we had preregistered. Participants in the no-time-pressure condition gave slower estimates ($M_{no\ time\ pressure} = 9.94$, $SD_{no\ time\ pressure} = 0.50$, $N_{no\ time\ pressure} = 79$) than participants in the time-pressure condition ($M_{time\ pressure} = 9.74$, $SD_{time\ pressure} = 0.56$, $N_{time\ pressure} = 91$), $t(167.94) = 2.47$, $p = .007$ (one-tailed), $d = 0.378$.

3.2.3. Hypothesis test

A 2 (time pressure: no vs. yes) \times 2 (anchor: low vs. high) ANOVA revealed no main effect of the anchor, $F(1, 166) = 1.04$, $p = .310$, $f = 0.079$. Time pressure did not have an effect on the estimate, either, $F(1, 166) = 0.06$, $p = .810$, $f = 0.019$. Whereas we hypothesized a Time Pressure \times Anchor interaction, no such effect was present, $F(1,$

Table 1

Descriptive Statistics for the Probability Estimate used in the Hypothesis Test.

Anchor	Time pressure	<i>M</i>	<i>SD</i>	<i>N</i>
Low (10)	Absent	17.82	17.82	44
High (90)	Absent	22.17	21.12	35
Low (10)	Present	17.96	20.92	53
High (90)	Present	20.21	22.60	38

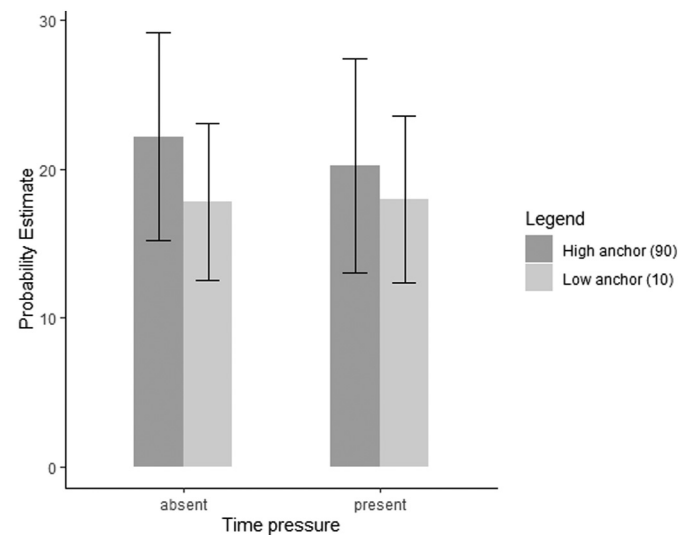


Fig. 3. Nonsignificant interaction between time pressure and subliminal anchor for estimating the probability of a lung pestilence in India. Note. Error bars represent 95% confidence intervals.

166) = 0.11, $p = .743$, $f = 0.025$, 95% CI = [0, 0.155]. This interaction effect was significantly smaller than the original effect of $f = 0.867$. Contrary to the original effect, post hoc analyses revealed that the anchoring effect was descriptively stronger in the no-time-pressure condition ($d_{no\ time\ pressure} = 0.226$, 95% CI = [-0.221, 0.672], $p = .333$, two-tailed vs. $d_{time\ pressure} = 0.104$, 95% CI = [-0.314, 0.520], $p = .631$, two-tailed). The data for the hypothesis test are presented in Table 1 and Fig. 3.

3.2.4. Exploratory analyses

All further analyses were not preregistered and were computed for exploratory purposes only. Beware of the high probability of false positives and false negatives given the high number of exploratory analyses and the relatively small sample size.

We checked whether participants in the online version had been less concentrated by comparing the scores of the letter task between the data from the lab and the online data. The mean letter task score was higher for the participants in the lab ($M_{lab} = 13.61$, $SD_{lab} = 1.40$, $N_{lab} = 111$) than for the online participants ($M_{online} = 13.52$, $SD_{online} = 1.30$, $N_{online} = 59$), but the difference was not significant, $t(126.06) = 0.40$, $p = .687$ (two-tailed), $d = 0.061$. Raw and log-transformed response times for the probability estimates and demographics did not differ between the participants in the lab and the online participants (all $ps > 0.109$, see <https://osf.io/jsg65/> for exact results).

We expanded the ANOVA reported for the hypothesis by the online versus offline factor in order to determine whether the interaction effect was stronger for the offline version than for the online version. The three-way interaction of anchor (low vs. high), time pressure (no vs. yes), and experiment version (offline vs. online) was not significant, $F(1, 162) = 2.08$, $p = .151$, $f = .110$. Conducting the letter task analysis and the expanded ANOVA for student status revealed only a main effect of the probability estimate, that is, nonstudents gave higher estimates on average.

To test whether the letter task influenced the strength of the

anchoring effect, we correlated the absolute difference between the estimate and the anchor with the letter task score. The correlation was almost zero, $r(168) = 0.015$, $p = .845$ (two-tailed).

3.3. Discussion

In our high-powered, preregistered, close replication of Reitsma-van Rooijen et al. (2006), we could not replicate the subliminal anchoring effect that they reported for participants who had to provide their estimates quickly. If anything, the pattern we found was actually the opposite of the hypothesized one because descriptively, the anchoring effect was stronger in the condition where time pressure was absent. Moreover, we found that the time-pressure manipulation significantly decreased response times for the probability estimate, which is a manipulation check that was not provided in the original study. We can alleviate concerns about the quality of the data as we were able to show that participants completed the tasks conscientiously (e.g., very high scores in the letter task) even in the online sample.

Two issues have to be noted, however. First, unlike Reitsma-van Rooijen et al. (2006), we did not conduct a pilot study to check that participants were unable to perceive the anchor. Although we would not expect differences between Dutch and German people with respect to their perception thresholds, our sample included nonstudents and was older. We found no correlation between age and the score in the prime detection task, which is why we do not view this deviation as a disadvantage. Again, note that the analyses yielded no differences between the offline student sample and the online mixed sample.

The more severe issue might be the fact that we deviated from our preregistration by ending recruitment in our lab and switching to an online version of the experiment. This put the (lack of) incentive back in line with the original study, however (in the original study, no incentive was reported, whereas participants in our lab were given candy). Because the experiment was hosted online anyway (and the participants in the lab sat in front of a computer screen), none of the code had to be changed. For the online sample, we were unable to check for whether the computer screen's refresh rates were at least 60 Hz, however, nor could we cover any numbers presented in the individual rooms where participants took the experiment. Although it is highly recommended (Plant, 2016), we could not test whether the presentation time matched the determined time. Moreover, online recruitment meant that nonstudents could take part in the experiment. In the original study, only students from Leiden University participated. Although subliminal priming has generally shown to be possible in using web-browser technology (e.g., Crump, McDonnell, & Gureckis, 2013, Study 7) the subliminal priming effects could be smaller in the online sample.

As our sample size was approximately three times the original sample size, and as the sample size of the online version was as large as the original study, we computed exploratory analyses to determine that the location of the experiment and the student or nonstudent status did not matter for outcome of the hypothesis test.

4. General discussion

In two high-powered, preregistered, close replications, we found no evidence for subliminal anchoring. The quality of the data was determined to be high, and thus, the failed replication attempts cannot be blamed on suboptimal data. We openly addressed all deviations from the original studies as well as from our preregistration, and we conducted exploratory

analyses to ensure that they had no relevant impact. Results from the three original studies and our replications are presented in Table 2.

4.1. Subliminal anchoring and incidental environmental anchoring

Subliminal anchoring can be considered similar to another specific kind of anchoring: incidental environmental anchoring (IEA; Critcher & Gilovich, 2008, p. 242). In IEA (e.g., Brewer & Chapman, 2002; Critcher & Gilovich, 2008; Wilson et al., 1996), anchors are presented as numbers that are part of the target stimulus such as the likelihood of a “fictitious college linebacker registering a sack in the conference playoff game” with jersey number 54 in the low anchor condition (vs. 94 in the high anchor condition; Critcher & Gilovich, 2008, Study 1). These two specific kinds of anchoring are similar in that they do not include the classical comparative question of anchoring (“Is it more or less than X?”), and for both, priming has been suggested as an explanation (Brewer & Chapman, 2002, p. 76, for IEA; Mussweiler & Englich, 2005, p. 135, for subliminal anchoring). Interestingly, IEA could not be replicated either (e.g., Edmonds, 2017; Klein et al., 2018), and previous results have been interpreted as a possible result of questionable research practices (Shanks, Barbieri-Hermitte, & Vadillo, 2020). However, the recent failures to replicate IEA (Klein et al., 2018; Shanks et al., 2020) did not attempt to address the replicability of subliminal anchoring.

It is important to mention that these two specific forms of anchoring are not linked theoretically, nor has previous research identified them as comparable, and they both lack a theoretical basis as they have been studied in a rather exploratory fashion. For example, Critcher and Gilovich (2008) stressed that they did not aim to “distinguish between the different theoretical accounts in the anchoring literature,” but they instead intended to “document the existence” of IEA (p. 243). Differences between IEA and subliminal anchoring could thereby have been crucial for replication, which is why we consider the present replication attempt necessary. For example, Mussweiler and Englich (2005) argued that the repeated presentation used in subliminal anchoring (10 times over 60 s in their studies) but not in IEA may “increase anchor accessibility above the critical threshold” (Mussweiler & Englich, 2005, p. 141). Note that in our studies, the anchors were presented in the center of the screen 10 times in Study 1 and 15 times in Study 2. Another difference that could have been crucial was participants' awareness of the source of the primed information, which is given in IEA but not in subliminal anchoring. “If people are aware of the source of the primed information, they are more likely to try to avoid being influenced by it, which often results in an overcorrection” (Wilson et al., 1996, p. 400). In line with this, it seems possible that IEA does not work due to respondents' awareness of the anchor - but subliminal anchoring could have effects.

The fact that both of these special kinds of anchoring could not be replicated now corroborates the assumption that deliberately considering the anchor as a potential value is essential (see also Mochon & Frederick, 2013) for anchoring to occur.

4.2. The possibility of false-positive effects in the original studies

The two failed replication attempts give rise to the question of why the original studies found effects. With only three reported p -values, there is too few data for a reliable p -curve analysis (Simonsohn, Nelson, & Simmons, 2014, p. 544; see also Carter, Schönbrodt, Gervais, & Hilgard, 2019). However, a look at the three exact p -values reveals that the original findings are not indicative of a very robust effect: The values reported by Mussweiler

Table 2
Sample Sizes and Cohen's d s for the Subliminal Anchoring Effect in the Original and Replication Studies.

Study	Time pressure	Original N	Original d	Replication N	Replication d
M&E 2005, Studies 1 & 2	Absent	76	0.690	165	-0.148
R&D 2006	Absent	30 or 32	-0.28	79	0.226
R&D 2006	Present	32 or 30	0.867	91	0.104

Note. Cell sample sizes for R&D 2006 are $N = 30$ for one group and $N = 32$ for the other group, but the study did not report which N corresponded to which group.

and English (2005) are $p = .041$ (Study 1, p. 136) and $p = .050$ (Study 2, p. 137). Reitsma-van Rooijen et al. (2006) reported $p = .025$ (p. 384) for the interaction effect between anchor (high vs. low) and time pressure (absent vs. present). Obtaining multiple p -values in this range is extremely unlikely both under the H_0 (i.e., when p -values are uniformly distributed) and under the H_1 (i.e., when the distribution of p -values is right-skewed; Simonsohn et al., 2014). Thus, these results might very well be due to chance or questionable research practices such as selective reporting.

4.3. Is the effect of subliminal anchors undetectably different from zero?²

The original studies were likely to have been false-positives and we did not find an effect in our replication studies. Does this mean that, our results suggest that the subliminal anchoring effect is undetectably different from zero? A method to answer this question is the small-telescopes approach proposed by Simonsohn (2015). The logic behind this approach is that we can conclude that there is no evidence for the phenomenon if we have taken a closer look (e.g., in the analogy of astrophysicists searching for planets, we must have “used a bigger telescope” in our study) than previous research – but still did not find anything. In that approach, the null-hypothesis is accepted if the replication effect size is significantly smaller than “a small effect”. A small effect is computed as the effect size for which the original study would have had only 33% power (i.e., $d_{33\%}$; Simonsohn, 2015, p. 4).

In the case of subliminal anchoring, we computed $d_{33\%}$ on the basis of Mussweiler and English (2005, Studies 1 and 2) because their sample size exceeds the one from Reitsma-van Rooijen et al. (2006) and because the pooled effect size of the experimental conditions of Reitsma-van Rooijen et al. (2006) is not significantly different from zero. The small effect that we derived on the basis of the small telescopes framework and Studies 1 and 2 from Mussweiler and English (2005) was $d_{33\%} = 0.250$, $N = 76$. The confidence interval of our pooled subliminal anchoring effect was $d_{95\% CI} = [-0.210, 0.224]$ and thus significantly different from $d_{33\%}$ (see Fig. 4). Therefore, we conclude that there is no evidence for the existence of subliminal anchoring. Note that these analyses were exploratory, they were added as a result of a reviewer's suggestion. We recommend future research to make a priori sample size considerations on the basis of the small telescope approach. Given our sample, we cannot exclude that subliminal anchoring could lead to very small effects. Minimum effect sizes for subliminal anchoring given our pooled sample size ($N_1 = 180$, $N_2 = 153$, one-tailed test) are $d_{95\%} = 0.362$ and $d_{80\%} = 0.274$.

4.4. Limitations

The most severe limitations of our replication studies are the deviations from the original studies. Although our replications were as close as possible to the original studies, this does not mean that there were no deviations. While designing our studies, we noted that Reitsma-van Rooijen et al. (2006) reported many more details than Mussweiler and English (2005). Closeness could thus be established more easily for the second replication study. Note moreover that the original materials were no longer available. Guaranteeing closeness was furthermore impeded by ambiguous or missing descriptions. For example, Mussweiler and English (2005) did not clarify what they meant when they stated that participants “indicated their suspicion about the flickering of the letter string,” nor did they identify the size of the letters on the screen. Reitsma-van Rooijen et al. (2006), on the other hand, had not described what exercise they used to familiarize participants with the letter task. In both cases, the closeness between our replications and the original studies is impossible to determine. A final factor limiting the closeness of replication was the Covid-19 pandemic that forced us to close our labs and continue our second replication as an online study although it had been designed as a laboratory study. Only 111 of 160 participants in study 2 (replication of Reitsma-van Rooijen et al., 2006) conducted the study in a controlled laboratory environment, the rest

participated online. This was still about two times the sample size of the original study and there were no differences between the online and the offline sample (e.g., the letter task scores were almost identical, response times did not differ) which makes us confident that form of participation did not make a difference.

In a recent discussion about replicator's degrees of freedom (Yeager, Bryan, & O'Brien, 2019), some replicators have been accused of exploiting researcher degrees of freedom (Wicherts et al., 2016) to produce non-significant results. We used the replication recipe preregistration template (Brandt et al., 2014) and sought to ensure that we did not exploit researcher degrees of freedom. This included methods of data analysis. The analytic choices had been made prior to data analysis, and the models were not complex (t -tests and ANOVA). The analysis code for the experiments and R scripts were written before we conducted the experiments. All changes in the code are highlighted. We furthermore highlighted all exploratory analyses, such as the small-telescopes approach. We recommend future research to take this approach into consideration for the determination of the necessary sample size.

4.5. The future of (Subliminal) anchoring

Instead of moving forward and building on existing bricks, we resolved a contradiction between two findings by showing that neither of them could be replicated. There is currently no convincing evidence of the existence of subliminal anchoring.

We believe that replication studies are necessary for the progress of anchoring research: We provide evidence that points to questionable research practices in subliminal anchoring research, which dovetails with findings by Shanks et al. (2020) regarding incidental environmental anchoring. Although we do not doubt the existence of classical anchoring effects, we advise researchers to use p -curve analyses and direct replications (e.g., Simonsohn et al., 2014) to test for the robustness of specific kinds of anchoring or moderators of anchoring. Accordingly, we advise future research to make use of preregistration (e.g., by using the template provided by van 't Veer & Giner-Sorolla, 2016; for examples of preregistered anchoring studies see Frech et al., 2020; Lewis et al., 2019). The non-replicability of anchoring effects in some paradigms might not be obvious to researchers when they try to create new models. Thus, the identification of non-replicable findings and the resolution of contradictions in the literature are necessary for the new models not to be wrong a priori.

Open practices

- All materials including the, analysis scripts, (translated) questionnaire files, stimulus generation code, and the data sets are available online at <https://osf.io/96de2/>. All analyses have either been preregistered or marked as exploratory. We highlighted all deviations from the preregistration and the analysis script. All preregistrations and all studies that we conducted on subliminal anchoring were reported. We encourage other researchers to use our data and materials or to conduct different replication studies to clarify the contexts in which subliminal anchoring can occur.
- Preregistration of Study 1: <https://osf.io/8urxv>
- Postcompletion registration of Study 1: <https://osf.io/acykd>
- Preregistration of Study 2: <https://osf.io/qxwhv>
- Postcompletion registration of Study 2: <https://osf.io/gh52k>

Contributions

- The first author developed the research questions, drafted the analysis scripts, and programmed the studies.
- The first and second authors oversaw the project and drafted the manuscript.
- All authors designed and conducted Study 1 and the respective preregistration.
- All authors revised and approved the final manuscript.

² We thank an anonymous reviewer for pointing us to the approach.

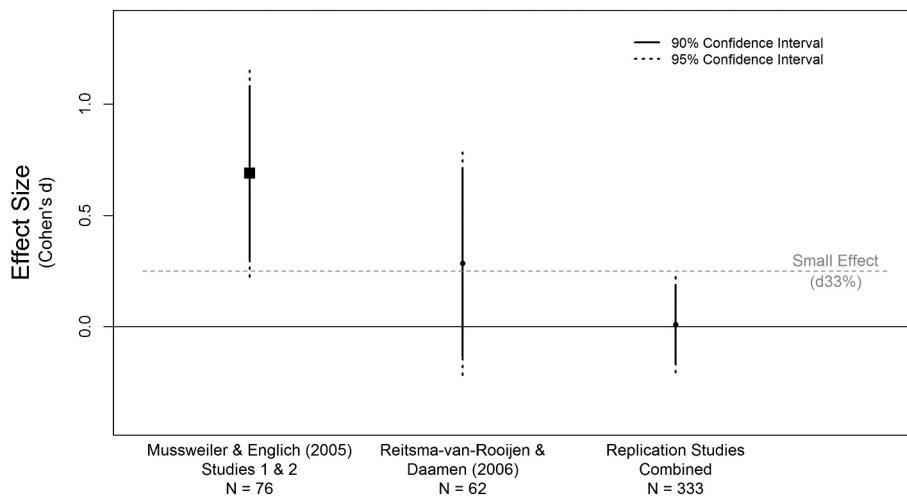


Fig. 4. Small telescopes analysis of subliminal anchoring effects.

Note. $d_{33\%}$ is based on the sample size of Mussweiler and English (2005). Results from Reitsma-van Rooijen et al.'s (2006) time pressure conditions have been pooled. This plot is based on the code provided by Simonsohn (2015; see <https://osf.io/qz5gw/> for the modified code).

Acknowledgements

This research was funded by a graduate scholarship granted by the state of Saxony-Anhalt, Germany, to Lukas Röseler. The funding source had no involvement in the study design, data collection, data analysis, interpretation of data, writing, or decision to submit.

We want to thank Frank Lammers and Annabel Heller for their help in recruiting participants, Jane Zagorski for language editing, and two anonymous reviewers for valuable suggestions on an earlier draft of the manuscript.

The reported studies were preregistered and all data and materials are available online (<https://osf.io/96de2/>).

References

- Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The sticky anchor hypothesis: Ego depletion increases susceptibility to situational cues. *Journal of Behavioral Decision Making*, 87(1), 23. <https://doi.org/10.1002/bdm.2022>.
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>.
- Brewer, N. T., & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making*, 15(1), 65–77. <https://doi.org/10.1002/bdm.403>.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115–144. <https://doi.org/10.1177/2515245919847196>.
- Chaxel, A.-S. (2014). The impact of procedural priming of selective accessibility on self-generated and experimenter-provided anchors. *Journal of Experimental Social Psychology*, 50, 45–51. <https://doi.org/10.1016/j.jesp.2013.09.005>.
- Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, 21(3), 241–251. <https://doi.org/10.1002/bdm.586>.
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), Article e57410. <https://doi.org/10.1371/journal.pone.0057410>.
- Edmonds, K. A. (2017). *Subliminal anchoring effects on weight estimates: Numerals versus numerosities*. Advance online publication <https://doi.org/10.13140/RG.2.2.17238.24649>.
- Elgendy, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., & Cichocki, A. (2018). Subliminal priming-state of the art and future perspectives. *Behavioral Sciences (Basel, Switzerland)*, 8(6), <https://doi.org/10.3390/bs8060054>.
- Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a within-subject, repeated-measures Ego-depletion paradigm. *Social Psychology*, 49(5), 271–286. <https://doi.org/10.1027/1864-9335/a000348>.
- Frech, M.-L., Loschelder, D. D., & Friese, M. (2020). How attribution-of-competence and scale-granularity explain the anchor precision effect in negotiations and estimations. *Social Cognition*, 38(1), 40–61. <https://doi.org/10.1521/soco.2020.38.1.40>.
- Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, 141(1), 124–133. <https://doi.org/10.1037/a0024006>.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.soc.2010.10.008>.
- Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., Urlichich, D., & Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000644> Advance online publication.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Michael, 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., ... Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>.
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., Babalola, M. T., Bahník, Š., Berkics, M., Bernstein, M., Jason, J. M., Berry, D. R., Bialobrzeska, O., Bocian, K., Brandt, M., Busching, R., Cai, H., Cambier, F., ... Nosek, B. A. (2018). *Many Labs 2: Investigating variation in replicability across sample and setting*. Advance online publication <https://doi.org/10.31234/osf.io/9654g>.
- Lewis, J., Gaertig, C., & Simmons, J. P. (2019). Extremeness aversion is a cause of anchoring. *Psychological Science*, 30(2), 159–173. <https://doi.org/10.1177/0956797618799305>.
- Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, 122(1), 69–79. <https://doi.org/10.1016/j.obhdp.2013.04.002>.
- Mussweiler, T., & English, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes*, 98(2), 133–143. <https://doi.org/10.1016/j.obhdp.2004.12.002>.
- Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgemental anchoring. *European Review of Social Psychology*, 10(1), 135–167. <https://doi.org/10.1080/14792779943000044>.
- Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136–164. <https://doi.org/10.1037/e413782005-562>.
- Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: An open letter. *Behavior Research Methods*, 48(1), 408–411. <https://doi.org/10.3758/s13428-015-0577-0>.
- Reitsma-van Rooijen, M., Daamen, D. D., & L. (2006). Subliminal anchoring: The effects of subliminally presented numbers on probability estimates. *Journal of Experimental Social Psychology*, 42(3), 380–387. <https://doi.org/10.1016/j.jesp.2005.05.001>.
- Röseler, J., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce judgment adjustment for both internally and externally generated anchors? *Journal of Experimental Social Psychology*, 87, 103942. <https://doi.org/10.1016/j.jesp.2019.103942>.
- Shanks, D. R., Barbieri-Hermitte, P., & Vadillo, M. A. (2020). Do incidental environmental anchors bias consumers' price estimations? *Collabra: Psychology*, 6(1), 19. <https://doi.org/10.1525/collabra.310>.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2160588> Advance online publication.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559–569. <https://doi.org/10.1177/0956797614567341>.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. <https://doi.org/10.1037/a0033242>.
- van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2–12. <https://doi.org/10.1016/j.jesp.2016.03.004>.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.)*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 7, 1832. <https://doi.org/10.3389/fpsyg.2016.01832>.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4), 387–402. <https://doi.org/10.1037/0096-3445.125.4.387>.
- Yeager, D. S., Bryan, C., & O'Brien, J. (2019). Replicator degrees of freedom allow publication of misleading 'Failures to Replicate'. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3408200> Advance online publication.