Meta-Psychology, 2025, vol 9, MP.2024.4137
https://doi.org/10.15626/MP.2024.4137
Article type: File-Drawer Report

Open data: Yes
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: Yes

Edited by: Matt N Williams
Reviewed by: Konstantinos Ioannidis, Weicong Lyu
Analysis reproduced by: Anica Scholz
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/KVZN5

Falsifying the Insufficient Adjustment Model: No Evidence for Unidirectional Adjustment from Anchors

Lukas Röseler[1,2], Lisa Incerti[1,3], Tobias Rebholz[4], Christian Seida[1], and Frank Papenmeier[4]
[1]University of Bamberg
[2]Münster Center for Open Science University of Münster
[3]University of Erlangen-Nuremberg
[4]University of Tübingen

After considering a more or less random number (i.e., an *anchor*), people's subsequent estimates are biased toward that number. Such anchoring phenomena have been explained via an adjustment process that ends too early. We present a formalized version of the insufficient adjustment model, which captures the idea that decreasing the time that people have to adjust from anchors draws their estimates closer to the anchors. In four independent studies (N = 898), we could not confirm this effect of time on anchoring. Moreover, anchoring effects vanished in the two studies that deviated from classical paradigms by using a visual scale or a two-alternative forced-choice paradigm to allow faster responses. Although we propose that the current version of the insufficient adjustment model should be discarded, we believe that adjustment models hold the most potential for the future of anchoring research, and we make suggestions for what these might look like.

Keywords: anchoring, TOTE, time pressure, two-alternative forced-choice, insufficient adjustment

When frogs want to cross streets, they often do not make it because many of them get killed by cars before they can reach their destination, which is the other side of the road (e.g., Gryz and Krauze, 2008). Although not as fatal and not exogenous, a similar problem has been identified for humans' numeric judgment. That is, when humans make judgments where they begin with a particular number, they often do not arrive at their desired target but stop too early. This phenomenon has been termed *anchoring* (or *anchoring and adjustment*; Tversky and Kahneman, 1974) and— although this process does not lead to immediate death—random or uninformative starting points have been shown to bias judges' decisions (Englich et al., 2006), experts' house price estimates (Northcraft and Neale, 1987), numerous negotiation outcomes (Bystranowski et al., 2021; Orr and Guthrie, 2006), willingness to pay (Li et al., 2021), or any numerical estimates (Röseler and Schütz, 2022). Theoretical explanations of anchoring have not been as solid, and numerous different accounts have come (e.g., Epley and Gilovich, 2001; Frederick et al., 2010) and gone (e.g., Bahník, 2021; Harris et al., 2019; Shanks et al., 2020). The earliest account and also the model that has faced the smallest number of contradictions and replication failures is the insufficient adjustment model, which is best described by the analogy of a frog trying to reach its destination but stopping too early. In this article, we conduct a strict test of the insufficient adjustment model.

Insufficient Adjustment Model

Tversky and Kahneman (1974) suggested that anchoring occurs when people begin a process of adjustment in numerical judgments with an experimenter-provided or self-generated number, adjust it in a certain direction, but then end the adjustment too soon — much like when a frog attempts to cross a road unharmed but is not successful. Thus, numerical estimates are biased toward previously considered numbers. Anchoring has commonly been shown in paradigms that ask participants to make numerical estimates after receiving a more or less random number (e.g., by spinning a wheel of fortune or by writing down the last digits from their Social Security Number).

Later, Epley and Gilovich (2001) developed the insufficient adjustment model, which describes the adjustment process via *TOTE units* (Miller et al., 1986), which were originally developed to explain a wide range of human behavior. TOTE in this case represents an acronym for the sequence of four successive phases, namely, Testing, Operating, Testing, and Exiting (Eisenberger et al., 2005). In the context of anchoring, it means first *testing* whether the initial value one had in mind (i.e., the anchor) seems plausible. If it does not, the *operating* phase follows: A certain value is added to or subtracted from the value under consideration. Consequently, the next *testing* phase consists of a further plausibility check. As soon as this check is positive, the process is terminated, or in terms of TOTE, the loop is *exited* (see also Eisenberger et al., 2005; Röseler, 2021; Röseler et al., 2020). If adjustments are still necessary after the second testing phase, the first two stages can be repeated as often as necessary until the value seems plausible. The result of the adjustment process then represents the estimated value, regardless of the number of runs. Epley and Gilovich (2001, 2006) explained that people tend to stop the adjustment process once they reach a value that seems plausible. As plausible values form a range around the actual value, anchored estimates are still too far away.

For a few years, the insufficient adjustment model was used only for effects that occurred with self-generated anchors, such as "0°C" for the freezing point of Vodka (Epley and Gilovich, 2001, 2004, 2005, 2006), whereas effects that occurred with experimenter-provided anchors were explained by the selective accessibility model via priming of anchor-consistent information and confirmatory hypothesis testing (Mussweiler and Strack, 1999a, 1999b). However, the distinction has been dropped (Chaxel, 2014; Simmons et al., 2010), and the selective accessibility model has been falsified (Bahník, 2021; Harris et al., 2019). Moreover, failed attempts to replicate results involving moderating variables have been published, casting doubt on the insufficient adjustment model altogether (Röseler, Bögler, et al., 2022).

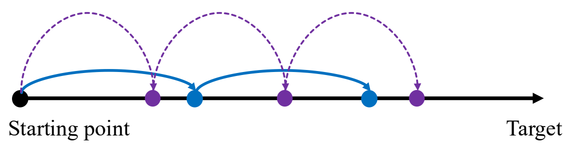
Apart from findings on moderating variables, most reports have made the insufficient adjustment model seem extremely plausible. For example, people actually report engaging in a stepwise adjustment process that is exclusively consistent with the insufficient adjustment model: Frech et al. (2020) created a paradigm in which participants indicated all their adjustment steps, which revealed a unidirectional adjustment process with a decreasing slope (see also Röseler et al., 2020). Note that Frech et al. (2020) used a price negotiation scenario, whereas Röseler et al. (2020) instructed participants to adjust unidirectionally to estimate the weights of animals.

The Time Hypothesis of Anchoring

Although TOTE units have been suggested to offer a potential way to formalize the insufficient adjustment model, little effort has been put into it (see also Röseler et al., 2020). Consider two frogs trying to cross a road.

Figure 1

Frog Analogy of TOTE Loops as a Model for Anchoring



Note. The distance covered from the starting point is the product of strength and stamina. Frog A (white arrow heads) has more stamina (three jumps) than Frog B (black arrow heads, two jumps), but Frog B's jumps are stronger and thus farther.

Both of them have limited strength, but they differ on numerous variables, such as how far and how often they jump. Frog A engages in many short jumps, whereas Frog B takes a few wide jumps. But both stop jumping before they reach their target because they are exhausted and need rest (see Figure 1). Characterizing the frogs with respect to their *strength* (average distance of a jump) and *stamina* (total number of jumps) provides a more fine-grained description of each of the frogs than the description that would be available from just looking at the final result (which is the total distance covered; or in anchoring terms, the final result is the *adjustment*).

The frog analogy is just a small step in formalizing the TOTE model. Looking at the adjustment that has been made away from the anchor over time, many more parameters can be added: The first and most obvious parameter when anchoring is viewed as a process is the parameter of time; that is, the total amount of adjustment depends on the time that has passed. More specifically, the anchor's influence is *stronger* if there is little time to adjust. Past research has tested the hypothesis that the time that people take to make judgments is associated with the strength of anchoring (Chaxel, 2014, p. 48). Chaxel (2014) did not find a correlation but note that this hypothesis cannot be derived from the TOTE model presented here if the parameters are assumed to be partly independent: If time is unlimited, individual differences in the process that result from differences in strength and stamina could very well even out over time. Figure 2 illustrates two adjustment functions and their variation in strength, speed (i.e., length of pauses), and number as proposed by the TOTE model. For example, in this model, we assume that the strength of the jumps is constant, that is, the last jump is as strong as the first.[1] The adjustment function that is depicted in Equation 1 results from this model, such that t represents the time that has passed since the adjustment process began, and *pause* represents the length of each pause. The Gaussian brackets around t/pause lead to discrete jumps and can easily be removed to create a continuous adjustment function.

$$\text{adjustment}(t) = \begin{cases} \left\lfloor \frac{t}{\text{pause}} \right\rfloor \cdot \text{strength}, & \text{if } t < \text{number} \times \text{pause} \\ \text{number} \cdot \text{strength}, & \text{if } t \geq \text{number} \times \text{pause} \end{cases}$$

We propose that this model should replace the insufficient adjustment account of anchoring due to the current model's greater empirical content. In this general form, the TOTE model has great potential and many advantages over other models: First, it can explain all kinds of adjustment processes (e.g., confirmation biases, Nickerson, 1998; emotional contagion and the inclusion-exclusion model, Schwarz and Bless, 1992; or adaptation level theory, Helson, 1948; emotion regulation, Scheier et al., 1994). Second, the parameters involved in these processes can be context-independent, suggesting that this is a macrotheory: For example, if a person's speed is measured as extraordinarily high in one paradigm, all processes that can be described by this model can be influenced by this speed. Third, it offers new, potentially more reliable ways to measure susceptibility to anchoring effects in comparison with classical methods (for a discussion, see Röseler, Weber, et al., 2022). Fourth, it allows and inspires research on moderators. For example, to better understand the specific paradigms, we suggest that commonsense correlates be looked for. The speed parameter might be strongly correlated with people's processing speed. Fifth, the adjustment process is proposed to occur in every situation for every person. Note that, although we do not investigate the specific model parameters (strength, number, pause) in this study, this type of model allows us to derive the time hypothesis, that is, with more time comes less anchoring.

Evidence For and Against the Time Hypothesis

In fact, there are five publications in which the time hypothesis of anchoring was tested.

- Yik et al. (2019) had participants infer the affect of protagonists in stories and put some of the participants under time pressure (i.e., they had to complete the entire task within a fixed amount of time). Anchors were not numerical but semantical. Anchoring effects in the time pressure condition were stronger than in the control group (Study 3, p. 5).
- Zong and Guo (2022) had consumers judge prices under time pressure or not (p. 5, H7) with a similar time pressure manipulation. Anchors were numerical (i.e., prices). They did not find a difference between the two groups (p. 11).
- Reitsma-van Rooijen and Daamen (2006) found that subliminally presented numbers anchored price estimates under time pressure but not when time pressure was absent. However, the subliminal anchoring effects that occurred with the time pressure manipulation could not be replicated in a later study (Röseler et al., 2021).
- Lieder et al. argued (2018a) and showed (2018b) that when people have high *time costs* when asked to give estimates, they rely on numerical anchors more strongly. However, their paradigm strongly deviated from typical anchoring paradigms, thus rendering it difficult to determine whether these effects can actually be called anchoring.
- Another relevant finding is the potential relationships between cognitive ability and susceptibility to anchoring. For example, Bergman, Ellingsen et al. (2010) found associations between cognitive reflection and anchoring strength. However, manipulation or measurement of time was not part of these experiments and they should be taken with great caution due to the reliability problem of anchoring discussed by Röseler et al. (2022).

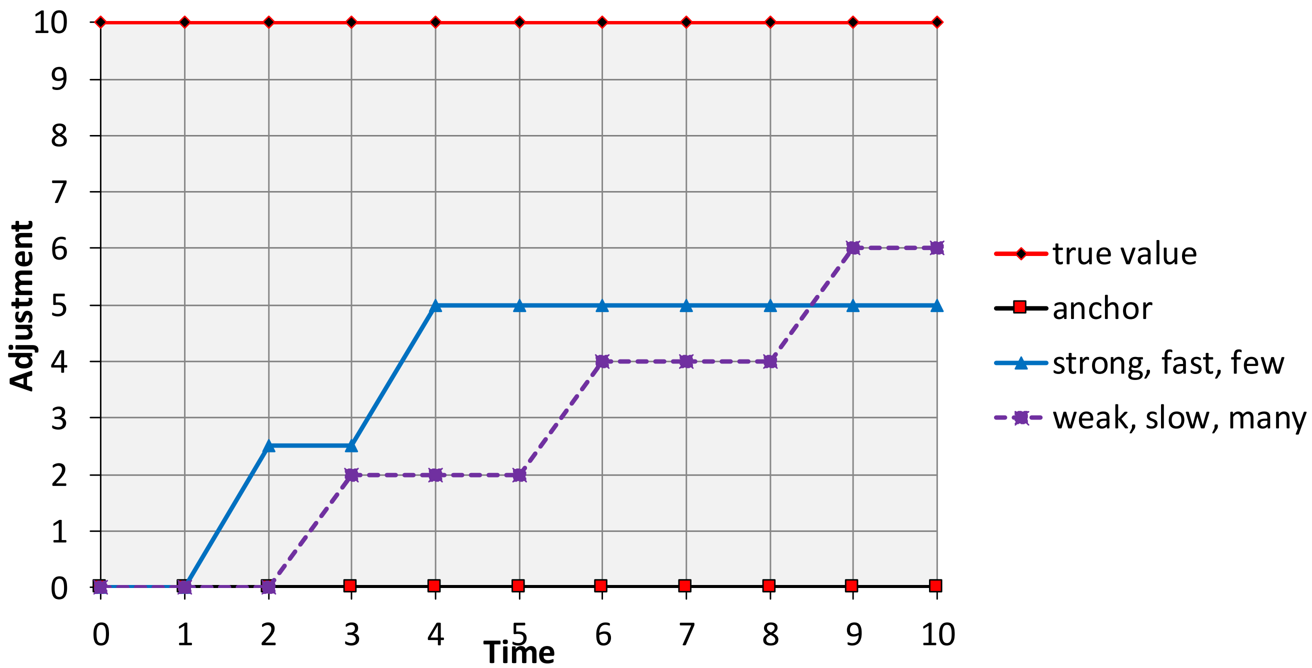
Taken together, the evidence is mixed but also not very informative. For example, individual studies could have gotten the unit of time wrong, as the model does not specify a priori whether adjustment occurs within (milli-)seconds or minutes. Also, some time pressure manipulations could have just led participants to adjust more quickly rather than stopping earlier, and for one of the studies, (subliminal) anchoring effects could not even be replicated, let alone their association with time pressure. Finally, despite the fact that most of the studies were conducted fairly recently, none of them were preregistered. Preregistered studies are less susceptible to questionable research practices (e.g., Brodeur et al., 2024) such as selective reporting (e.g., Kvarven et al., 2020).

To submit the time pressure hypothesis to a strict test, we conducted four preregistered studies with varying

[1] Our analyses of Frech et al.'s data actually showed that adjustment strength decreases with the number of steps (https://osf.io/ck4y3).

Figure 2

Detailed TOTE Model Incorporating Strength, Speed, and Number of Jumps Over Time



paradigms, items, and principal investigators. Although the unit-of-time argument can never be completely refuted, the heterogeneity across our study designs makes it much less plausible (see Table 1).

Study 1: River Lengths

Method

To test whether time pressure in the form of limited time to give estimates affects reliance on the anchors that are provided, we conducted a preregistered (https://osf.io/c2ynz) study in which time pressure was manipulated between-subjects by giving participants either an unlimited or a limited amount of time to estimate the lengths of 30 rivers. The data were collected from November to December 2021. We recruited participants via mailing lists, collected data online via SoSciSurvey (Leiner, 2019), and analyzed the data with R Version 4.2.1 (R Core Team, 2018) and the packages pwr (Champely, 2020), psych (Revelle, 2018), and ggplot2 (Wickham, 2016).

Materials

Anchoring items were taken from Schultze and Loschelder's (2020) study on advice taking (p. 3, Table 1) and consisted of 30 rivers for which participants had to estimate the length in kilometers via an open numeric response field. We conducted a pretest to check for anchoring effects. Anchors were twice the true value (high anchor) or the rounded half of the value (low anchor). While participants in the main study received all 30 items, we chose the six items with the strongest anchoring effects (Rio Grande, Rio Purus, Tejo, Volga, Murrumbidgee, Sambesi) and put them on top of the list of items. We assigned high and low anchors to them such that the mean river lengths, anchors, and effect sizes were approximately equal (see https://osf.io/qrpce for the exact values of all items). Therefore, for the first six items, participants received low and high anchors in an A-A-B-B-B-A-design. In other words, half of all participants received high anchors for Items 3 through 5 and low anchors for the other items and vice versa. Anchors were presented as hints, disclosing that the true value is below or above the anchor. Hypothesis tests were focused on only the first six items. Participants were not aware that all the other presented items were dropped from our hypothesis tests. This procedure was used so that participants in the time pressure condition had enough time to answer the items while still experiencing time pressure.

The maximum amount of time that people could take in the time pressure condition was determined by using

Table 1

Overview of Study Designs and Results Regarding the Presence of Anchoring Effects (AEs) and Time Effects (TEs)

S	PI	Paradigm	Anchoring items	Manipulation	Location	AE	TE
1	LR	Classical anchoring	River lengths	Limited total time	Online	Yes	No
2	CS & LR	Visual response scale	City distances	Limited time per item	Online	Mixed	No
3	LI & LR	2AFC	Number of "b"s vs. "d"s in matrix	Limited time per item	Lab	No	No
4	FP & TRR	Classical anchoring	Classical items (Jacowitz & Kahneman, 1995)	Limited time per item	Online	Yes	No

Note. S: Study. PI: Principal investigators (letters in this column are abbreviations of the authors' names). 2AFC: two-alternative forced-choice. AE: Anchoring effect present? TE: Time effect present? Anchors were experimenter-provided in all studies and participants did not receive rewards for accuracy. For quantitative results and a mini meta-analysis, see Figure 17.

data from the pretest. The median time was 254 s (M = 355 s, N = 31), and the fastest participant took 152 s. We limited the time to 120 s via the questionnaire software, which seemed sufficient to estimate the first six river lengths and which was still less than the time the fastest participant took. During the task, participants in the time pressure condition saw the number of seconds that remained in bold at the top of the questionnaire page.

Procedure

Participants were informed that they had to be at least 18 years old and that the anonymized data set would be published. They were asked to answer all questions in the presented order. After being assigned to an anchor condition (see Materials) and a time pressure condition (time pressure vs. no time pressure), some of them were instructed that they should try to answer all 30 items within 2 min.

Finally, participants answered three time pressure manipulation check items (I hurried up, I felt time pressure, I considered the answers in detail [reversed]), one exploratory item (My estimates leaned toward the hints), and four items that we later applied to exclude participants from the analyses (I conducted the study on a computer, I have taken part in this study more than once, I looked up the answers [e.g., via a search engine], I know the psychological phenomenon called the "anchoring effect"). After completing the study and regardless of their responses to the additional questions, participants were offered immediate feedback on their estimates, the true values, and how close their estimates were relative to all other participants via a web application.

A Priori Sample Size Determination

We expected the time effect to be a difference of d = 0.5 in the anchoring effects between the two conditions (e.g., d = 1.0 in the control condition and d = 1.5 in the time pressure condition). Therefore, we collected data from N = 180 participants to achieve 95% power for a one-tailed two-sample t test. This target sample size was also sufficient for anchoring effects: The mean anchoring effect size for the six central items on the pretest was d = 1.79. As we chose the items with the largest effect sizes, we expected the effect size to regress to the mean (e.g., d = 1). Simultaneously, our design allowed us to conduct a within-subjects test of anchoring, that is, to compare mean estimates for the high-anchor items with those for the low-anchor items (see https://osf.io/2rkbn for the power analysis code).

Analysis Plan

As per our preregistration, we excluded participants if at least one of the following criteria applied: (a) did not finish the survey (i.e., did not answer the manipulation check items), (b) reported to have looked up the correct answers for the anchoring task, (c) took part in the study more than once, (d) had missing values on at least one of the first six anchoring task items, or (e) did not conduct the study on a computer as requested in the instructions.

To conduct a manipulation check for the anchoring items, we computed the mean of the three items unless their internal consistency was $\alpha < .70$, in which case we planned to use only the first item.

Within-subjects susceptibility to anchoring effects was computed by subtracting the mean estimates for the three low-anchor items from the mean estimates for the three high-anchor items. Therefore, larger scores reflected stronger susceptibility to anchoring. We expected these scores to be larger in the time pressure condition when tested with a one-tailed *t* test for independent samples. The entire analysis script was preregistered, and changes or additions are marked as "CHANGED" or "ADDED," respectively (https://osf.io/2rkbn).

Results

Our sample consisted of 180 participants (92 in the time pressure condition; 96 women, 83 men, 1 diverse). Their ages ranged from 20 to 79, with a median age of 55. This sample size is sufficient to detect effects of $d > 0.492$ with 95% power. Whereas almost all the participants in the control condition responded to all the items, most of the participants in the time pressure condition completed only 12-17 items.

Data Quality Checks

We tested for anchoring effects on an item-by-item basis and found large effects for all six items (see Figure 3). The average within-subjects anchoring effect (see Analysis Plan) was very large, too. On average, participants in the high-anchor condition gave estimates that were 1566.85 km longer than those in the low-anchor condition despite the finding that the average river lengths were the same, $t(179) = 15.80$, $p < .001$ (one-tailed), 95% CI [1402.86 km, ∞ km], $dz = 1.177$.

The internal consistency of the manipulation check items was poor, $\alpha = .485$, 3 items, $N = 180$), so we used only the first item for our manipulation check as we had preregistered. Experienced time pressure was not larger in the time pressure group than in the control group, $t(175.35) = -0.34$, $p = .367$ (one-tailed), $d = 0.05$, 95% CI [-0.24, 0.34], and mean time pressure was only slightly larger in the time pressure group ($M_{time\ pressure} = 3.88$, $SD = 1.22$, $N = 88$; $M_{control} = 3.82$, $SD = 1.13$, $N = 92$).

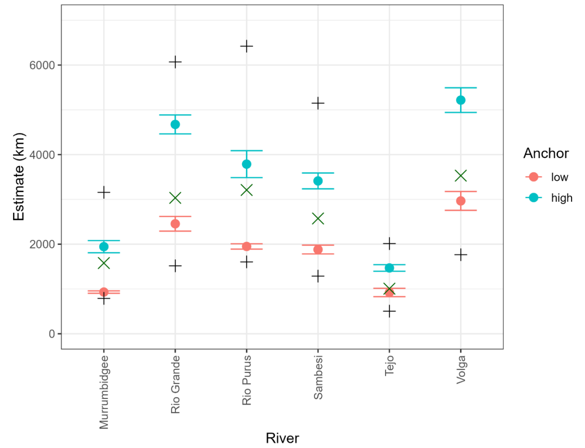
Participants in the time pressure condition responded more quickly than those in the control condition ($Med_{time\ pressure} = 124$ s, $Med_{control} = 266$, $w = 7921$, $p < .001$ (one-tailed). Due to imprecision in the time measurement from the online survey tool, the time limits were not actually 120 s but were as high as 148 s.

Hypothesis Tests

The strength of the anchoring effects in the time pressure condition was not significantly larger than in the control condition ($M_{time\ pressure} = 1629.42$ km, $SD = 1299.71$ km, $N = 88$; $M_{control} = 1507.00$ km, $SD = 1364.07$ km, $N = 92$), $t(178) = 0.62$, $p = .269$ (one-tailed), $d = 0.09$, 95% CI [-0.20, 0.38] (see Figure 4).

Figure 3

Anchoring Effects for the Six Central Items



Note. For each item, + represents low and high anchors, × represents the true value (i.e., river length in km), dots with error bars represent mean estimates with 95% confidence intervals for the low and high anchor groups.

Exploratory Analyses

We conducted the manipulation check as planned, that is, due to the low internal consistency of the three items, we used only the first item (see Data Quality Checks). However, when we used the other two items, people reported a statistically significant increase in time pressure in the time pressure group. We furthermore tested whether any of the single manipulation check items was related to susceptibility to anchoring, which was not the case. The results for all the items, including the mean of all items, are presented in Table 2.

Discussion

We asked participants to estimate 30 river lengths and induced time pressure for half of them by limiting the total available time to less time than anyone from the pretest needed to answer the items. Although this time limit decreased the median duration of the task from 266 to 124 s, participants did not report greater time pressure on the central manipulation check item but only on the two other items. The anchoring effects were very large for all the items, but the time limit did not lead to stronger anchoring effects. As discussed above, when considered alone, this failure to corroborate the time hypothesis is of limited value, as the time pressure may simply have not been strong enough, or it could have led people to adjust more quickly but not to stop earlier.

Table 2

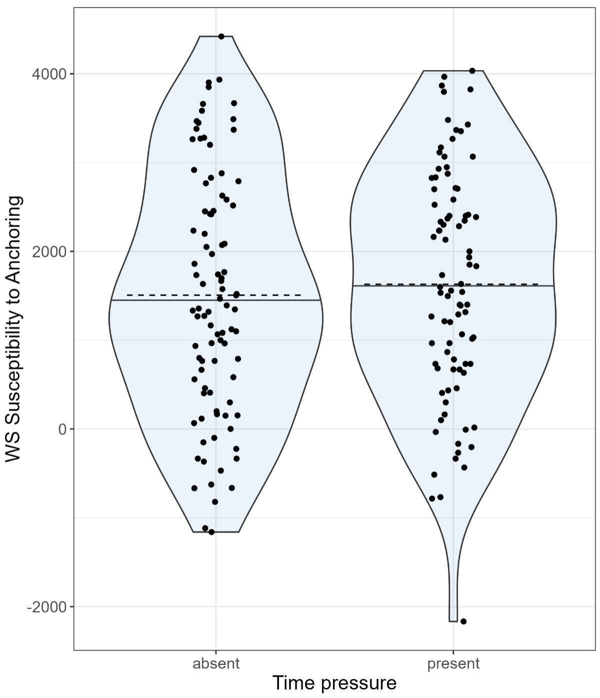
Effect Sizes for Manipulation Check Items

Manipulation check item	Cohen's *d* [95% CI]	*r* [95% CI]
I hurried up.	0.05 [-0.24, 0.34]	-.053 [-.198 , .094]
I felt time pressure.	1.53 [1.20, 1.87]	-.045 [-.190, .102]
I considered the answers in detail (reversed).	0.32 [0.02, 0.61]	.100 [-.047, .242]
Mean score	0.94 [0.63, 1.25]	-.013 [-.159, .134]

Note. Positive *d* values indicate the experience of higher time pressure in the time pressure condition. *r* is the correlation between the respective item and susceptibility to anchoring.

Figure 4

Within-Subjects Susceptibility to Anchoring in the Time Pressure and Control Groups



Note. WS: within-subjects. Solid lines represent medians, and dashed lines represent means. Values larger than 0 indicate an anchoring effect across the six items. Values are unstandardized within-subjects mean differences between the high and low anchor items. That is, on average, a person with a value of 1,500 estimated the three high-anchor river lengths to be 1,500 km more than the three low-anchor river lengths.

Deviations From the Preregistration

We deviated from the preregistration in one way; that is, we corrected a coding error in the preregistered analysis script that would have prevented the manipulation check item from being determined. After the correction, the analysis script functioned as described in the script's comments and as described in the corresponding preregistration.

Study 2: City Distances

Method

To test whether time pressure in the form of limited time to give estimates affects reliance on provided anchors, we conducted a preregistered (https://osf.io/56vte) study in which time pressure was manipulated within-subjects by giving participants different amounts of time (4, 10, 16, or 22 s) to estimate distances between 12 pairs of cities. This study's main feature was a trial-by-trial time limit similar to Lieder et al.'s (2018b) study design. As most people are not able to type a four-digit number into an open response field in 4 s, we replaced the open response field commonly used in anchoring research with a 580-pixel-wide horizontal visual response scale that included labeled end points and a mark and an arrow that represented the anchor. The data were collected from February to March 2021.

We recruited the participants from the student body, collected data online via SoSciSurvey (Leiner, 2019), and analyzed the data with R Version 4.2.1 (Team, 2018) and the packages psych (Revelle, 2018), ggplot2 (Wickham, 2016), tidyverse (Wickham et al., 2019), lmerTest (Kuznetsova et al., 2017), BSDA (Arnold and Evans, 2021), and doBy (Højsgaard and Halekoh, 2018).

Materials

Participants estimated distances between 12 pairs of European cities on a visual response scale (min = 0 km, max = 1,300 km) in a fully randomized 2 (type of anchor: high vs. low) × 6 (true distance between pair of cities) within-subjects design. Thereby, each of the four time limits (time: 4 vs. 10 vs. 16 vs. 22 s) was applied three times. Anchors were computed by adding or subtracting to or from each true flight distance 0.4 times its own value (for the R code, see https://osf.io/2m7yj).

The 12 pairs of cities consisted of six pairs of city pairs with two of the pairs of cities always sharing the same true distance between them. In each pair of city pairs, one had a high anchor and one had a low anchor. We determined that there should be six high and six low anchors per participant. A list of all cities and their flight distances is available online (https://osf.io/eqwn3). Time limits were made salient by presenting an animated bar that decreased in length as the amount of remaining time decreased. To encourage participants to use the time in the cases in which it was larger (e.g., 24 s), we prevented them from making estimates by clicking on the scale until only 2 s were left. During the last 2 s, the response scale turned from grey to black. A sample item is presented in Figure 5. The code for reproducing the anchors (https://osf.io/2m7yj) and the bars from which the animated GIF files were created (https://osf.io/uqprj) are available online.

Procedure

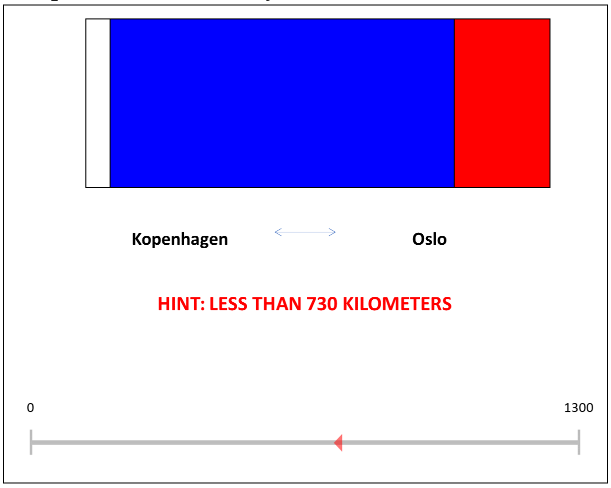
Participants were instructed that the purpose of the study was geographical knowledge and that there would be a quiz that they would take. We clarified that participation was voluntary and anonymous and was possible only via a computer. After they completed the study, they could receive immediate feedback about their estimates, the true values, and the current sample's mean estimates via an online Shiny App. Moreover, they had the chance to win one of two 10€ vouchers.

Participants had to acquaint themselves with the visual response scale in seven practice trials. In the first four trials, they had to mark different values on the scale (500, 250, 1000, 750). In two further trials, the time limit was introduced. Participants were shown animated bars that displayed the remaining time. Finally, a trial with the time limit and the anchor was presented. Afterwards, they completed the anchoring task, which consisted of 12 items.

After the anchoring task, participants could opt in to receive feedback, course credit, or the chance to win one of the two 10€ vouchers (or all of these options). Finally, we asked whether they thought of anchoring effects during the study.

Figure 5

Sample Item From Study 2



Note. Participants' task was to click on the grey line to indicate what they think the flight distance between the cities presented above is. All distances and anchors were between 411 and 866 km, with scale extrema fixed to 0 and 1,300 km for all pairs of cities. The red triangle on the grey line represents the anchor presented in the hint. The white, blue, and red sections of the bar above are indicators of the remaining time. The bar was entirely blue at the beginning of the task and shrank. When the blue bar was gone and the red bar began to shrink, the grey line turned black, and participants were able to click on the line until the entire bar was white. Participants were instructed about all paradigm features and completed multiple practice runs.

A Priori Sample Size Determination

As the effect size of time manipulations in anchoring paradigms was unknown at the time when the study was conducted, we assumed sufficient sensitivity for $N = 250$. An effect size of $dz \geq 0.209$ was thereby required for 95% power for a one-tailed paired t test (R code available online, https://osf.io/rmbxz).

Analysis Plan

As per our preregistration, we excluded participants if they changed their browser tab during the estimation task at least once, as this could indicate a low level of focus or that participants looked answers up. Moreover, participants were required to have clicked on larger values for the 1,000 km instruction item than for the 750 km instruction item to ensure they understood how to use the scale. To be included in the hypothesis test, participants were required to have at least two valid estimates per time condition.

Adjustment strength was standardized with the formula |[anchor]-[estimate]|/|[true value]-[anchor]|, where 1 corresponded to estimates of the true value and 0 to estimates of the anchor. To assess the quality of the data, we planned to check the number of missing values by time condition. The presence of anchoring effects was planned to be tested by comparing whether the mean adjustment scores were significantly smaller than 1 with a linear mixed effects model. We planned to test the time hypothesis using a one-tailed paired *t* test on the aggregated data set that compared mean adjustment strength in the 4 s condition with that of the 22 s condition. The analysis script was preregistered and is available online (https://osf.io/a3y4t).

Results

Data Quality Checks

Our sample consisted of 290 participants and 2,720 valid estimates. Using a linear mixed model with restricted maximum likelihood estimation with adjustment scores nested in participants and the corresponding *t* test using Satterthwaite's method, we tested whether participants' adjustment scores were significantly different from the true value, which was the case, $t(264.22) = 14.28$, $p < .001$ (two-tailed). However, they were not biased toward the anchor (< 1) but rather biased away from the anchor (> 1), $M = 1.29$, $SD = 0.88$, $N = 2{,}720$.

We tested whether the estimates were reliable (i.e., whether people who clicked on high values for one stimulus clicked on high values for other stimuli), which was not the case ($\alpha = 0.53$, average $r = 0.1$, 12 items). Unsurprisingly (Röseler, Weber, et al., 2022), the adjustment scores were not reliable either ($\alpha = 0.56$, average $r = 0.1$, 12 items).

As participants in the conditions with very little time might not have had enough time to give their estimates we tested how the number of missing values differed between time conditions. The number of nonmissing values in the 16 s condition was highest and in the 4 s condition was lowest. This result might be due to a lack of time for reading and comprehending the respective items in the 4 s condition rather than a lack of time for clicking (Figure 7).

Hypothesis Tests

We compared the 4 s condition with the 22 s condition in order to test the time hypothesis of anchoring, which is agnostic about the exact number of seconds needed to adjust from an anchor. Adjustment was similar between the two conditions, $t(498.62) = -0.51$, $p = .305$ (one-tailed), $d = -0.045$, 95% CI [-0.217, 0.128]. This result did not support the time hypothesis. For an overview of adjustment strength in all four time conditions, see Figure 8.

Exploratory Analyses

Using a saturated mixed effects model, we tested for the average effects of anchor, time, and true values with estimates nested within participants. Estimates were strongly correlated with true values, indicating that participants did not give random responses, $t(2633) = 2.68$, $p = .008$, $b = 0.39$. Higher anchors led to lower estimates, $t(2642) = -3.68$, $p < .001$, $b = -0.56$. Thus, increasing the anchor by 100 km decreased the estimates by approximately 56 km. Longer times did not affect the estimates, $t(2662) = 0.64$, $p = .521$, $b = 3.93$. Anchoring strength depended on the true values, $t(2638) = 2.46$, $p = .014$, $b < .001$. None of the other terms were significant. An overview of all estimates is provided in Figure 9 and Table 3.

Discussion

According to the time hypothesis, limited time should lead to stronger anchoring effects. To test this hypothesis with very small amounts of time (e.g., 4 s) while still granting participants enough time to give an estimate, we used a visual response scale, as it has been used in rare other cases (e.g., Lammers and Burgmer, 2017). The use of the visual response scale led the anchoring effects to become contrast effects; that is, high anchors led to lower estimates than low anchors. In other words, people adjusted too much from the anchors and shot over the true value in the opposite direction. Similar effects have been reported for anchoring effects (Röseler, Weber, et al., 2022; Sweeny et al., 2011) and might be due to an inherent link between assimilation (e.g., anchoring) and contrast (e.g., Müller-Lyer illusion) phenomena (Song et al., 2019). Recently, Spicer et al. (2022) discussed how nearby values can repulse (i.e., "push away") judgments, and distant values can attract (i.e., "pull toward") judgments (see also Bless and Schwarz, 2010; Rader et al., 2015). Spicer et al. (2022) also used visual response scales, and repulsion effects have not yet been reported in paradigms that used open numeric response fields. Apart from the visual response field, both our and Spicer et al.'s study highlighted the anchor on the scale. In an as yet unpublished study, we manipulated response scales to test whether visual response scales would increase the likelihood of repulsion effects, but the findings were inconclusive.

Although the time participants had to think about their responses was indeed different, the time pressure might have been conceived of as effectively the same across all the conditions. The reason is that responses

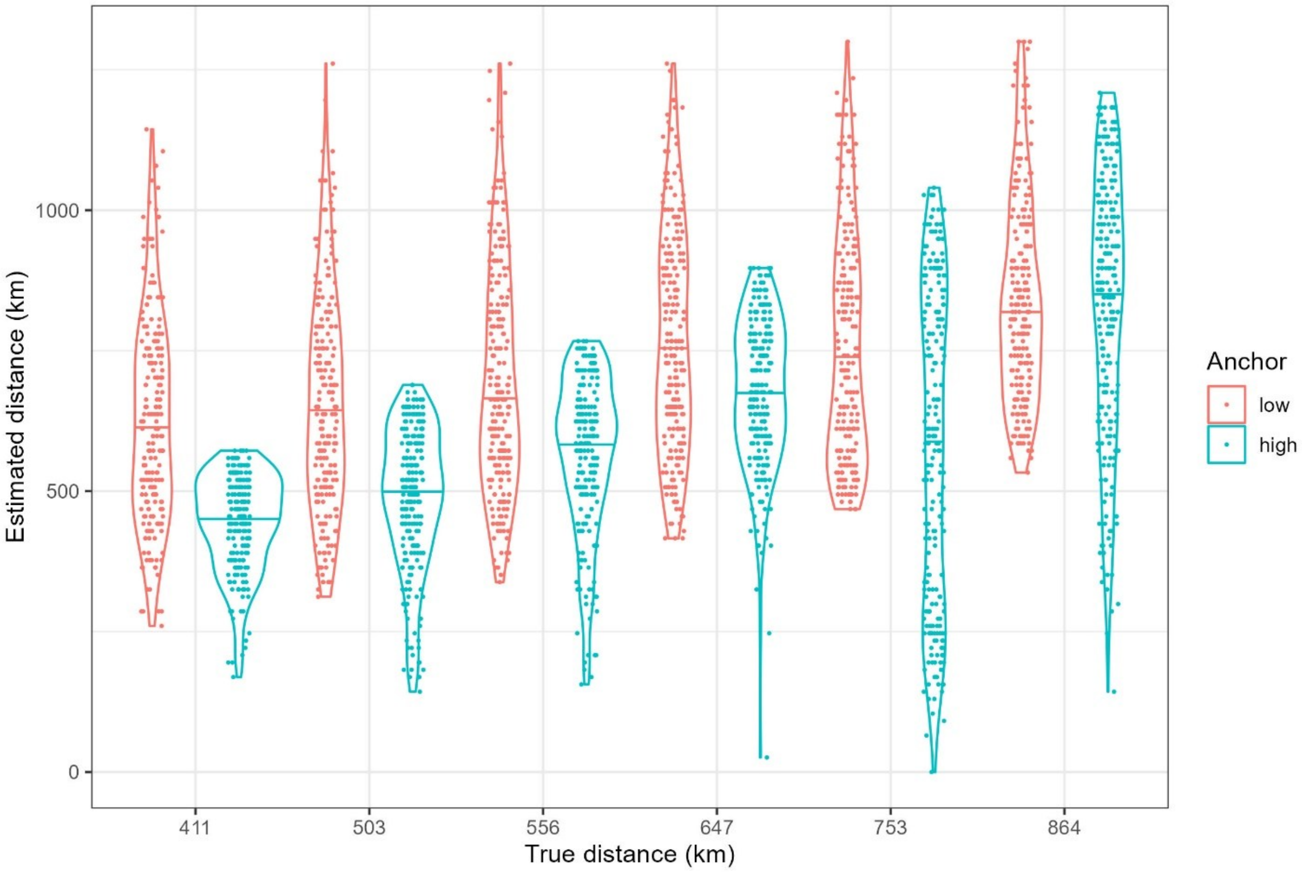
Table 3

City Distance Estimates for High and Low Anchors, Time Conditions, and True Values

Time Condition (s)	Anchor	True Value (km)	Mean	SD	N	95% Confidence Interval
4	low	411	619.07	179.54	29	[684.41, 553.72]
10	low	411	630.13	171.29	70	[670.26, 590]
16	low	411	608.60	193.79	65	[655.71, 561.49]
22	low	411	650.00	191.00	43	[707.09, 592.91]
4	high	411	414.96	85.82	25	[448.6, 381.32]
10	high	411	452.71	86.81	51	[476.53, 428.88]
16	high	411	443.12	90.52	81	[462.84, 423.41]
22	high	411	444.05	79.14	38	[469.21, 418.89]
4	low	503	683.28	201.85	25	[762.4, 604.16]
10	low	503	639.52	195.75	67	[686.39, 592.65]
16	low	503	680.88	191.70	56	[731.08, 630.67]
22	low	503	645.50	189.25	78	[687.5, 603.5]
4	high	503	518.70	115.71	40	[554.56, 482.84]
10	high	503	480.78	130.15	58	[514.27, 447.28]
16	high	503	460.98	120.46	63	[490.73, 431.24]
22	high	503	498.00	129.72	78	[526.79, 469.21]
4	low	556	682.75	189.41	79	[724.51, 640.98]
10	low	556	705.59	198.13	58	[756.58, 654.6]
16	low	556	666.71	201.22	49	[723.05, 610.37]
22	low	556	714.26	216.59	53	[772.58, 655.95]
4	high	556	579.68	147.06	44	[623.13, 536.23]
10	high	556	552.81	160.88	63	[592.54, 513.08]
16	high	556	569.76	122.73	58	[601.34, 538.17]
22	high	556	564.11	127.78	56	[597.57, 530.64]
4	low	647	685.13	195.75	47	[741.09, 629.16]
10	low	647	780.66	186.81	59	[828.33, 732.99]
16	low	647	770.00	198.55	65	[818.27, 721.73]
22	low	647	808.44	204.71	64	[858.59, 758.28]
4	high	647	651.03	122.93	38	[690.11, 611.94]
10	high	647	666.11	166.08	71	[704.74, 627.48]
16	high	647	684.86	115.64	69	[712.14, 657.57]
22	high	647	688.75	114.32	53	[719.53, 657.98]
4	low	753	707.78	154.40	36	[758.21, 657.34]
10	low	753	778.08	219.75	61	[833.23, 722.94]
16	low	753	785.97	222.34	61	[841.76, 730.17]
22	low	753	707.51	195.73	59	[757.45, 657.56]
4	high	753	596.22	261.03	51	[667.85, 524.58]
10	high	753	584.28	256.00	54	[652.56, 516]
16	high	753	605.22	303.03	72	[675.22, 535.23]
22	high	753	548.13	296.00	67	[619.01, 477.26]
4	low	864	758.33	168.41	36	[813.35, 703.32]
10	low	864	865.80	160.65	60	[906.45, 825.15]
16	low	864	833.55	187.89	67	[878.54, 788.56]
22	low	864	869.76	183.98	63	[915.19, 824.33]
4	high	864	731.44	267.19	34	[821.25, 641.63]
10	high	864	798.32	240.16	66	[856.26, 740.38]
16	high	864	892.90	222.65	73	[943.98, 841.83]
22	high	864	842.48	225.35	67	[896.44, 788.52]

Figure 6

City Distance Estimates for High and Low Anchors



Note. For each true value, there were two city pairs that had the same flight distance. Estimates were given on a visual response scale and could be between 0 and 1,300 km. Dots for each of the six different true distances are jittered to enhance readability.

had to be given within the last 2 s of any time pressure condition. In other words, the effective time pressure *for responding* did not differ between conditions. Accordingly, this might be an alternative explanation for the null effect in the study at hand. However, a greyed-out version of the response scale was present already as the time decreased, so that participants could think about their estimate and prepare their answer beforehand. Therefore, it is rather likely that the adjustment processes began before the response period.

Deviations From the Preregistration

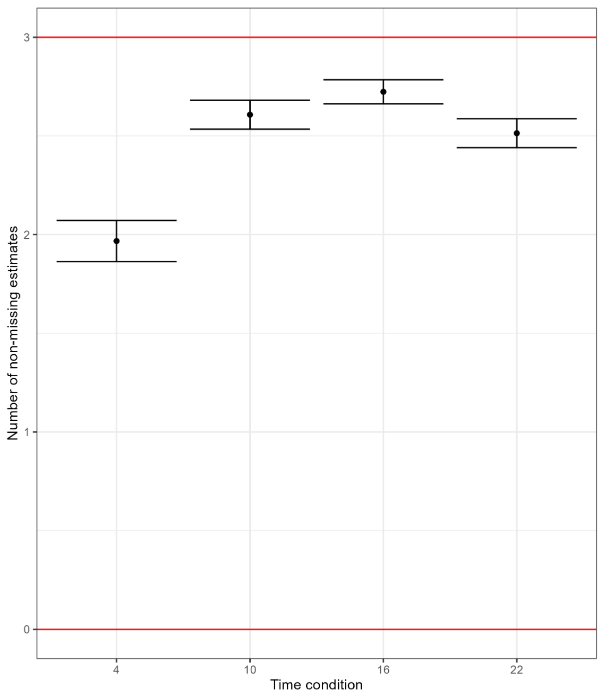
We deviated in one way from the preregistration: We corrected an error in the preregistered analysis script that involved converting from coded values into kilometers. This change and additions (e.g., plots) are marked in the analysis script (https://osf.io/rmbxz) as "CHANGED" and "ADDED," respectively.

Study 3: Two-Alternative Forced-Choice Paradigm

The main goal of this study was to implement a critical test of the time hypothesis, which is an investigation of whether anchoring effects are amplified under time pressure or, more fundamentally, whether, given an anchor, a distractor that is closer to the anchor than the true value is more preferred when additional time pressure is induced. In this case, however, we did not intend to use a standard anchoring paradigm as used in most previous study designs. That is, responses were not freely generated by the participants

Figure 7

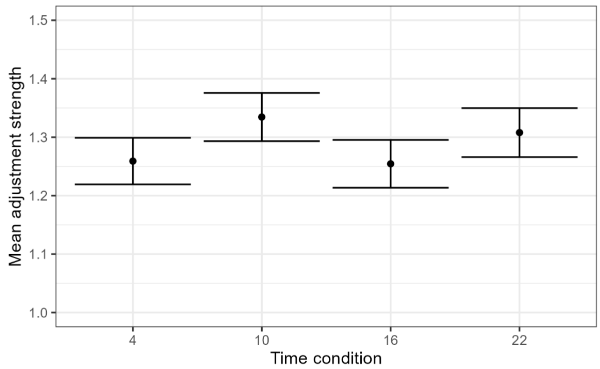
Number of Nonmissing Estimates per Time Condition



Note. For each time condition, participants could have between zero and three nonmissing values as indicated by the red lines. For example, a value of zero represents a person who gave no estimates within the given time in the respective time condition.

Figure 8

Mean Adjustment From Anchors by Time Condition



Note. The y-axis is truncated for better visibility.

but were entered via key presses in a within-subjects two-alternative forced-choice (2AFC) paradigm.

To our knowledge, there are no reports of anchoring effects being implemented in a 2AFC paradigm. We hypothesized that anchoring effects would occur when a 2AFC paradigm was used. In terms of the paradigm, when participants are given an anchor (in our study, a hint about the answer to a question) before choosing between a distractor (close to the anchor) and a true value (farther from the anchor), they should be more likely to choose the distractor than when they are not given the anchor.

Presuming that anchoring effects occur in this 2AFC paradigm, we further hypothesized that anchoring effects would be larger when participants had little time to respond.

Note that in this paradigm, high susceptibility to anchoring is confounded with highly inaccurate (i.e., incorrect) estimates. To exclude inaccuracy as an alternative explanation for less correct choices in the time pressure condition, we also included a baseline condition in which there was no anchor. For anchoring effects to be present, the number of incorrect responses should be larger in the anchoring condition than in the baseline condition, as anchors should bias judgments toward the distractor and away from the correct value. The time pressure effect in the condition without an anchor should be smaller than the time pressure effect in the condition with an anchor (see Figure 10). The data were collected in April and May 2022.

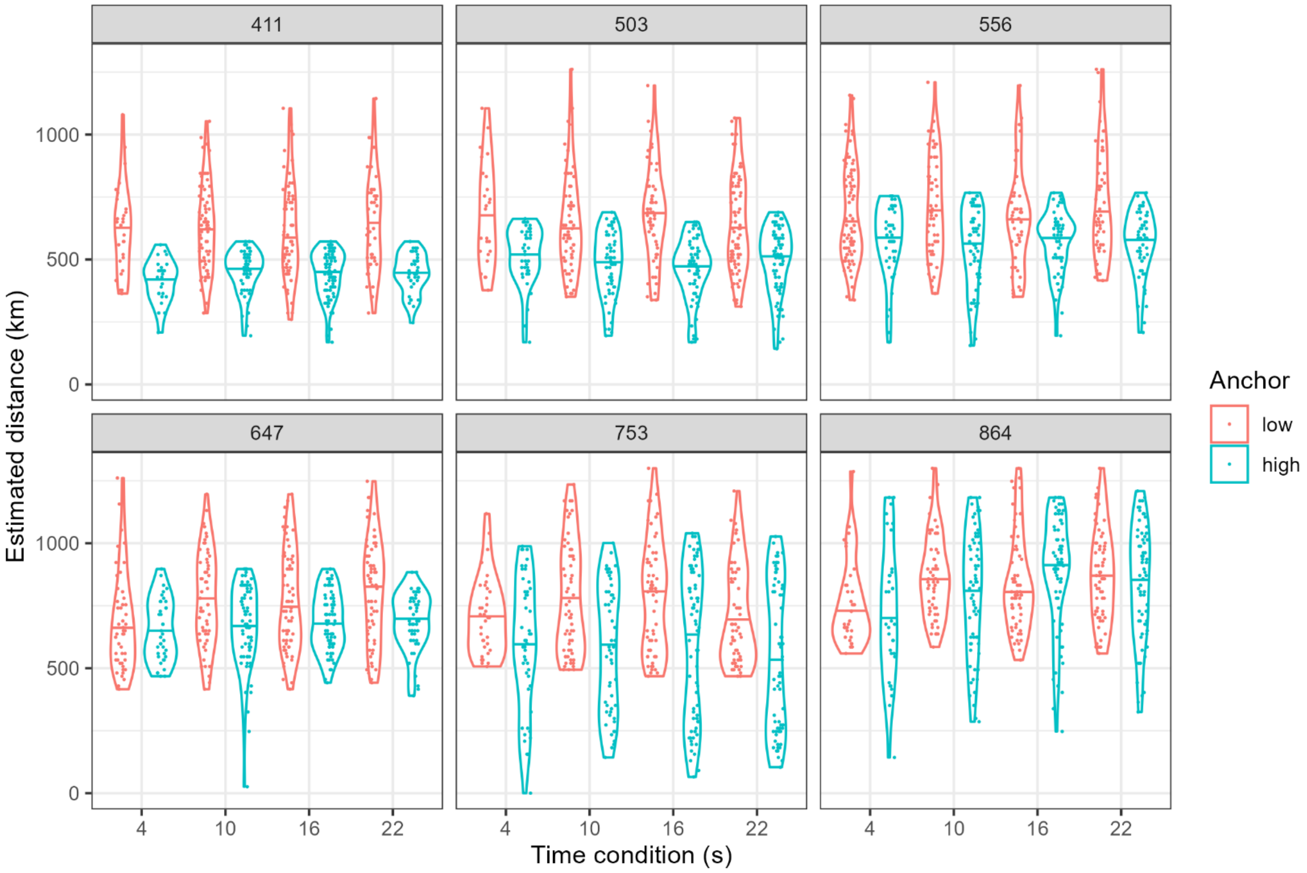
Note. The anchoring effect hypothesis is represented by the number of correct answers in the anchor condition (right side) being lower than in the no anchor condition. A larger number of correct answers is represented by the no time pressure condition (blue line) being above the time pressure condition (red line). The time pressure hypothesis of anchoring is represented by the difference between the no anchor and anchor conditions being larger for the time pressure condition than for the no time limit condition. We did not actually have a specific hypothesis about the main effect of time (i.e., the mean difference between the red and blue lines), but visualizing the hypothesis in an interaction plot is necessarily unambiguous.

Method

All available study materials and the data set can be accessed online (https://osf.io/8cwpy). In the present study, we used the software PsychoPy® (Peirce et al., 2019) to program and conduct the experiment and the software R Version 4.1.1 (Team, 2018) and the packages dplyr (Wickham et al., 2018), ggplot2 (Wickham, 2016), lmerTest (Kuznetsova et al., 2017), psych (Rev-

Figure 9

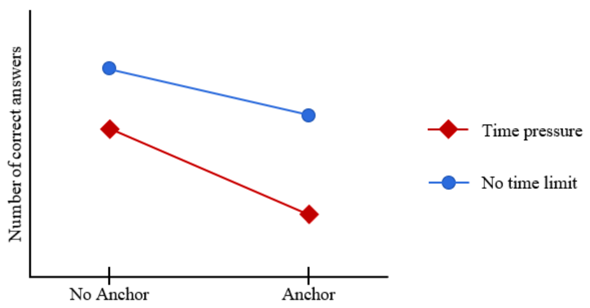
City Distance Estimates for High and Low Anchors, Time Conditions, and True Values



Note. The values in the grey boxes represent the true flight distances. For each true value, there were two city pairs that had the same flight distance. Estimates were given on a visual response scale and could be between 0 and 1,300 km. Time was manipulated from 4 to 22 s in steps of 4 s.

Figure 10

The Hypothesized Interaction and Main Effects of Anchor and Time Pressure in the 2AFC Paradigm



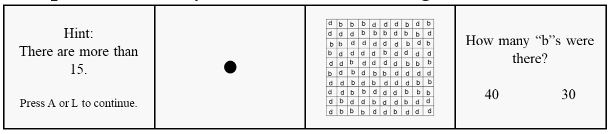
elle, 2018), lme4 (Bates et al., 2015), effects (Fox and Weisberg, n.d.), simr (Green and MacLeod, 2016), WebPower (Zhang and Mai, 2022), lsr (Navarro, 2015), and stringi (Gagolewski, 2022) to create letter matrices and finally analyze the data. The entire analysis script as well as the script for generating the letter matrices was preregistered in advance (https://osf.io/yuh78).

A Priori Sample Size Determination

Due to having no data to rely on with regard to effect sizes for anchoring or time in the newly created within-subjects 2AFC paradigm, we chose a target sample size of $N = 100$. In contrast to classical paradigms, our paradigm required participants to take part in a rel-

Figure 11

Sample Stimulus for the 2AFC Paradigm



Note. Original stimuli were presented with German instructions. The code to reproduce the stimuli is available online (https://osf.io/bs7f3). There are 40 “b”s in this matrix. Presentation time for the matrix was always 2 s, presentation time for the question varied.

atively lengthy study (i.e., 48 test and 16 target trials per participant instead of 1 - 1̃0 anchoring items). A power analysis revealed that the minimum effect size for the interaction between time pressure and anchor was a difference in Cohen's *d* in time pressure effects between the conditions with and without an anchor of 0.4, 0.5, and 0.75 for approximately 87%, 97%, and >99.9% power, respectively (for details and assumptions, see https://osf.io/eqnpb).

Materials and Procedure

The study was carried out in the University of Bamberg lab on laptops and desktop computers with PsychoPy® installed. First, participants were personally welcomed by the investigator and asked to enter the experimental room. They each sat down at one of the computers and began the experiment on their own. A cover story was used to disguise the true purpose of the study: Participants were told that the study was about estimation accuracy. They were asked to estimate the number of letter "b"s in a letter matrix that included the characters "b" and "d" by choosing one of two response alternatives (see Figure 11 for a sample stimulus). In addition, it was also explained that there was a variation in the response time (time pressure vs. no time limit) and either a hint or no hint. Then, age and gender as demographic variables were recorded.

Subsequently, the actual survey began with 16 training runs to ensure that the task had been understood. The transition to data collection was not marked, so participants were not aware that the first runs were training runs. In each case, participants were first told which conditions would be examined in the following block of tests (e.g., hint and time pressure). Then, each participant was presented with either a hint or no hint (e.g., "There are more than 15"). A jittered letter matrix with the letters "b" and "d" followed automatically for 2 s (see Figure 11). Participants were asked to choose between two alternatives by pressing a key ("A" for the left response and "L" for the right) to identify how many "b"s were contained in the matrix. In the time pressure condition, an elapsing time bar was displayed on the screen. When the time (1 s) was up and participants had not given a response, the screen turned red and participants were asked to hurry up ("FASTER!"). In the no time pressure condition, participants' response time was not limited. The test blocks were each randomized within the training block and within the survey block. In addition, the trials were administered randomly within the test blocks. The study included a total of 64 trials, 16 per survey block.

After the 64 runs (training and survey blocks), participants were asked to answer a few more questions (about whether they had to hurry when the response time was limited, whether they completed the study on a computer[2], and whether they had taken part in the study more than once). These questions were designed to test whether the manipulation in the time pressure condition had worked: Participants were under time pressure and therefore could not think about their answers in detail. Subsequently, participants were informed about the actual background of the study, including information about how we were investigating whether anchoring effects could be demonstrated in a 2AFC paradigm and whether the effects would be amplified under time pressure. In addition, there was an explanation of what anchoring effects are.

Finally, participants had the opportunity to generate a personal code, which could be used to retrieve their own results from a website. Additionally, the code could be used to retrieve their course credit. Completing the entire study took about 15 min. Figure 12 provides an overview of the flow of the entire study.

Measures and Analysis Plan

For the calculations, the total number of correct answers was recorded as the dependent variable. Many different independent variables were incorporated into the study design. Table 4 provides an overview of all variables.

Our study was carried out with a fully randomized within-subjects design. We told participants that the study followed a two-by-two design containing the presence versus absence of a hint and the presence versus absence of time pressure. We did not tell them that there were variations in the true number of “b”s (40 vs. 60), anchor direction (high vs. low), anchor distance (25 vs. 35), distractor direction (if there was an anchor,

[2] We originally planned to run the study online, which would have allowed participants to complete it on a tablet or mobile phone. Due to technical problems, we had to run it in the lab.

Table 4

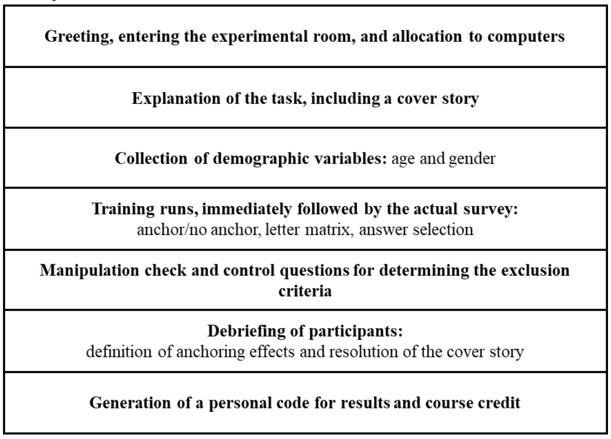
Overview of Independent variables

Factor	Levels			
Anchor	Present		Absent	
Time pressure	Present	Absent	Present	Absent
True value	40	60	40	60
Anchor direction	High	Low	No anchor	No anchor
Anchor distance	25	35	No anchor	No anchor
Distractor direction	Equal to anchor direction		High	Low
Distractor distance	10	20	10	20

Note. The different variable specifications were all combined with each other in our study. The list merely gives an overview of which variable expressions were included and varied within the anchor versus the no anchor conditions.

Figure 12

Study Procedure



Note. From the moment the participants split up to sit at the computers, all further explanations and tasks ran on the screens. The experimenter left the room after greeting.

it was equal to the anchor direction, otherwise high vs. low), and distractor distance (10 vs. 20). After all variants were combined, participants first went through 16 training runs before taking the actual survey with a total of 58 trials.

In the model we used in our analyses, we considered only the anchor and time pressure variables, thus corresponding to the two-by-two design that was also apparent to the participants. To test our first hypothesis, we calculated the main effect of anchor (absent vs. present). To test our second hypothesis, we computed the interaction between time pressure and anchor. The exact statistical methods we used to perform these calculations are described in the Results section.

Results

Participants

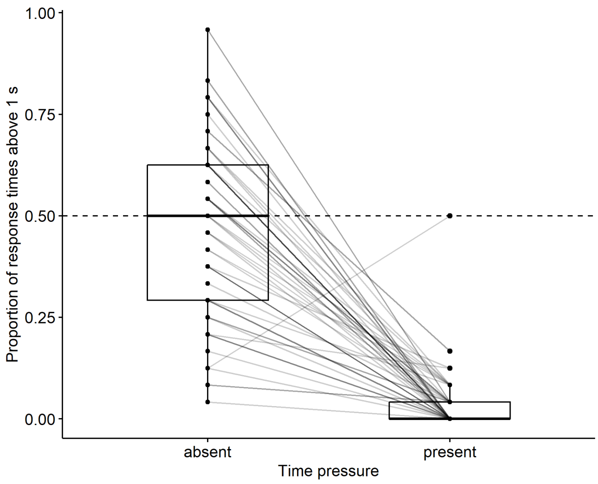
Because our original plan to conduct an online study could not be realized due to server problems suffered by the software provider, we had to switch to on-site data collection at the authors' university within a short period of time. Nevertheless, 60 participants (41 female, 8 male, 11 with no indication of gender) were recruited for the study. Their ages ranged from 18 to 56 years ($M = 25.97$, $SD = 8.10$, $N = 58$; two individuals did not provide age information). Participants were informed about the experiment via mailing lists from the University of Bamberg and via social networks. The study itself took place on site at the university. Therefore, mostly students from the University of Bamberg took part in the study. Regardless of their performance, they were offered course credit for their participation. Upon completion of the study, all participants had the opportunity to receive feedback on their results. To be eligible, participants had to speak German as their primary language. Due to the smaller-than-planned sample size ($N = 60$), power for Cohen's *d*s of 0.4 and 0.5 was only approximately 66% and 85%, respectively (instead of 87% and 97%, respectively, for $N = 100$; power analysis code available online, https://osf.io/eqnpb).

Data Quality Checks

To determine whether our time pressure manipulation had achieved its effect, we asked participants about the perceived impact of our time pressure condition at the end of the study. They were asked whether they had to hurry in this condition and whether they felt stressed. Finally, we used an inverted item to ask whether participants had been able to think in detail about their answers in the time pressure condition (four items, see Ta-

Figure 13

Paired Boxplot of Percentages of Responses Given After More Than 1 s for the No Time Pressure (Left) and Time Pressure (Right) Trials



Note. One of the participants gave exactly 50% of their answers after the warning in the time pressure condition. However, the 50% was not surpassed in this case either, which is why this participant's data were included in the analysis of the results.

ble 5; α = .49). The distribution of responses suggested that self-reported time pressure was high (see Table 5). Thus, the participants were not only noticeably more often faster than 1 s in their reaction times in the time pressure condition (see Figure 12), but they also had the subjective feeling that they were under time pressure and had to rush.

Hypothesis Tests

We ran a generalized linear model with a logit link function and judgments nested in participants to predict whether judgments were correct from the time-pressure condition and the anchor (Table 6). None of our predictors (i.e., the anchors themselves, the induction of time pressure, or the interaction between them) had a significant effect on whether participants chose the distractor that was close to the anchor or the true value, which was far away from it. The likelihood of correct responses was not associated with the time-pressure, $b = 0.179$, $se = 0.134$, $z = 1.339$, $p = .181$, or with the presence of anchors, $b = -0.105$, $se = 0.115$, $z = -0.915$, $p = .360$. Moreover, the two factors did not interact, $b = -0.266$, $se = 0.163$, $z = -1.630$, $p = .103$ (see Table 6). Thus, there is no evidence that anchoring occurred in the 2AFC paradigm or that the intensity of anchoring was moderated by time pressure (interaction).

Exploratory Analyses

We tested for effects of anchors and time-pressure with an additional exploratory model in which z-scored aggregated proportions of correct responses were predicted by time-pressure and anchor conditions. The model strongly converged with the preregistered model in that none of the effects of interest were significant, that is, there were no effect of anchoring, time-pressure, or the interaction (Table 7).

Descriptive statistics were calculated for each group to provide more detailed insights and to determine whether the results would show tendencies in the direction of our hypotheses. For this purpose, the various combinations of conditions were considered individually on the basis of their z-values. The condition that included time pressure and no anchor ($M = 0.30$, $SD = 1.17$) yielded the highest proportion of accurate responses, followed by the no anchor and no time pressure condition ($M = 0.04$, $SD = 1.05$). These findings are consistent with our idea that the presence of an anchor, meaning a hint in our study, would result in responses that were biased toward this hint. Although the finding was not significant, participants nevertheless chose the distractor less often in conditions without an anchor. The next best performance was achieved in the condition with an anchor and without time pressure ($M = -0.11$, $SD = 0.87$). The lowest scores were found in the condition with an anchor and time pressure ($M = -0.23$, $SD = 0.82$). Indeed, these observations were consistent with the hypotheses, but they did not reach statistical significance.

As our hypotheses assumed that the time pressure effect in the condition without an anchor would be larger than the time pressure effect in the condition with an anchor, we also tested this assumption with our data. For this purpose, we computed a paired t test on the aggregated proportions of correct responses, but again, the results were not significant, $t(59) = -1.66$, $p = 0.051$ (one-tailed), $d_z = 0.21$. To summarize, neither anchoring effects nor time effects could be detected in our paradigm. Nonetheless, the results showed tendencies in the direction of our hypotheses.

Discussion

We aimed to find out whether anchoring effects occur or can be detected in a newly created 2AFC paradigm. A critical test of the time hypothesis was carried out using a fully randomized, within-subjects design. Neither anchoring nor time pressure effects emerged in the paradigm. Compared with sample sizes from other studies, our study may have been underpowered. However, our use of a within-subjects design and our re-

Table 5

Response Frequencies of Intensity Levels to Manipulation Check Questions.

Questions	Intensity levels				
	Not at all	A little	Neutral	Much	Very strong
With the limited response time, I had to hurry.	3	5	5	22	25
The time bar made me feel stressed.	7	9	5	27	12
Even with the limited response time, I thought about my answers in detail.	10	35	8	7	0
I followed the hints.	6	25	13	15	1

Note. $N = 60$. In our study, the questions were presented to the participants in German. Internal consistency was poor, $\alpha = .49$.

Table 6

Fixed Effects of the Predictors and Their Interaction

Predictor	Estimate	*SE*	*z*	*p*
(Intercept)	0.38	0.10	3.73	<.001
Time pressure	0.18	0.13	1.34	.181
Anchor presence	-0.10	0.11	-0.91	.360
Time Pressure × Anchor interaction	-0.27	0.16	-1.63	.103

Note. SE = standard error; *df* = degrees of freedom. Dependent variable was whether judgments were correct.

cruitment of participants on site in a controlled environment should have compensated for the small sample size. Nevertheless, we encourage researchers to conduct studies to explore alternative anchoring paradigms more thoroughly to determine whether anchoring is paradigm-dependent (see also Frederick and Mochon, 2012; Mochon and Frederick, 2013).

Limitations and Future Research

Due to local limitations while conducting the study in person at the University of Bamberg, data could be collected from only 60 participants. As described in more detail above, we had actually intended a sample size of 100 participants for our study. This point is a clear deviation from our preregistration. A larger sample size or more trials would have increased the accuracy of the study and would increase the generalizability of the results. Future research should therefore consider conducting the study in an online format to be able to generate a larger number of participants and thus calculate more meaningful results. Another option would be to have a longer survey period at multiple sites if future studies want to employ a face-to-face format. In the process of examining a more extensive sample, future research should additionally focus on a more equal distribution in terms of the age, gender, and current employment of the participants in order to obtain a generalizable cross-section of the population for the survey.

Steps can also be taken to optimize the methodology. With the way our study was carried out, the adjustment processes—or more precisely, the exact steps—could not be recorded in detail. Future research should therefore consider using the think-aloud method (Ericsson and Simon, 1980; see also Epley and Gilovich, 2001, Study 1; Frech et al., 2020; Röseler et al., 2020) to be able to conduct a more detailed exploration of the leaps that might occur in participants' judgment process.

Deviations From the Preregistration

We deviated from the preregistration in one aspect, as we originally aimed to achieve a sample size of 100 participants but were able to collect data from only 60. All other aspects that had been planned in the preregistration were met accordingly.

Study 4: Classical Anchoring Items

Method

We conducted a preregistered (https://osf.io/yhrua) study to test whether time pressure in the form of limited time to give estimates would affect reliance on anchors that were provided for five classical items from

Table 7

Fixed Effects of the Predictors and Their Interaction

Predictor	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	0.04	0.13	209.79	0.35	0.728
Time pressure	0.25	0.16	177.00	1.58	0.117
Anchor presence	-0.15	0.16	177.00	-0.95	0.345
Time Pressure × Anchor interaction	-0.38	0.23	177.00	-1.67	0.096

Note. SE = standard error; *df* = degrees of freedom. Dependent variable was the z-scored aggregated proportions of correct responses by participant.

Jacowitz and Kahneman (1995). Time pressure was manipulated between-subjects by giving participants either an unlimited or a limited amount of time to provide their estimates. The experiment was run as part of a psychology course at the University of Tübingen with 110 students who were required to recruit five naïve participants (i.e., people who did not know they study's hypothesis) each. Data were collected online via SoSciSurvey (Leiner, 2019) from October 25, 2021 until November 4, 2021, and analyzed with R Version 4.1.2 (Team, 2018) and the packages sjstats (Lüdecke, 2018b), pwr (Champely, 2020), ggpubr (Kassambara, 2020), tidyr (Wickham and Girlich, 2022), dplyr (Wickham and Girlich, 2022), Hmisc (Harrel, 2020), jtools (Long, 2020), ggplot2 (Wickham, 2016), and effectsize (Ben-Shachar et al., 2020).

Materials

We used an experimental 2 (anchor: high vs. low) x 2 (time pressure: with vs. without) between-subjects design. Each participant responded to the five estimation tasks that were originally reported by Jacowitz and Kahneman (1995, p. 1163, Table 1) and are included in Table 8. We also used the same anchors as reported in Jacowitz and Kahneman (1995, p. 1163, Table 1), but because the items were presented in the German language, we transformed the anchors from feet to meters and from mph to km/h by rounding to the nearest whole number.

Procedure

Participants were invited to take part in the study using the students' or their own computer via a link to the online study. On the first page, we informed them that their participation was voluntary and anonymous and that their data would be made publicly available. To participate in the study, they had to give their consent and confirm that they were at least 18 years old. Participants were asked about their age and gender before they were provided with general instructions about the study procedure. Specifically, participants were informed that they would be asked to complete five estimation tasks for which they would also be provided with another participant's example answer[3] that would be visible for 5 s. On the following page, in both conditions (with and without time pressure), participants were asked not to use the Internet to find the right answer (see Materials). In the condition with time pressure, they were additionally instructed that they had at most 7 s to answer an item and click the continue button after the example answer had disappeared.

The two items listed first in Table 8 were included in the first two estimation tasks and were preregistered to be excluded from the analyses. They served as training trials, which were particularly necessary to prepare the participants who were under time pressure to be able to answer in time. The other three items were randomized across the remaining trials and included in the analyses. The transition from the training trials to data collection was not marked, so participants were not aware that the first two trials were training trials. Participants were provided with high or low anchors during the estimation phase depending on which anchoring condition they were assigned to. After the fifth and last item, participants were fully debriefed and asked for their informed consent to use their data.

A Priori Sample Size Determination

The study was designed to provide sufficient power to detect an interaction effect with an effect size similar to or larger than the interaction effect reported in Yik et al.'s (2019) Study 3, namely, η_p^2 = .020. Thus, we

[3] Note that while some studies of anchoring have used mechanisms that communicate the anchor as random (e.g., Ariely et al., 2003; Tversky and Kahneman, 1974), Jacowitz and Kahneman (1995) do not disclose whether they used a similar strategy. Many more recent studies present anchors as other participants' estimates or without additional information (e.g., Bahník, 2021; Frech et al., 2019; Lee and Morewedge, 2021).

Table 8

The Five Original Items From Jacowitz and Kahneman (1995) and Their German Translations Used in Study 4

Item number	Type of trial	Original question	German translation	Low anchor	High anchor	Preregistered outlier criteria
1	Training	Height of Mount Everest (in feet)	Wie hoch ist der Mount Everest?	610 m	13868 m	-
2	Training	Year telephone was invented	In welchem Jahr wurde das Telefon erfunden?	1850	1920	-
3	Test	Height of tallest redwood (in feet)	Wie hoch ist der größte Küstenmammutbaum der Welt?	20 m	168 m	[5, 500]
4	Test	Number of United Nations members	Wie viele Mitgliedsstaaten gehören den Vereinten Nationen an?	14	127	[10, 300]
5	Test	Maximum speed of house cat (in miles per hour)	Wie viel km/h beträgt die Höchstgeschwindigkeit einer Hauskatze?	11	48	[5, 80]

required a sample size of $N = 388$ participants (four groups of 97 participants each) in order to achieve a power of at least 80% to detect the predicted interaction effect for a significance level of 5% (Faul et al., 2007; see https://osf.io/c9gez).

Analysis Plan

We measured the estimates (numbers) given in response to the estimation tasks. Participants provided their estimates by typing them into the text input fields in the online study. Because our estimation tasks used different scales, we explained in the preregistration that we would transform the responses to each estimation task into z-scores prior to the analyses. Our main dependent measure for each participant was thus their mean z-score across the three estimation tasks that were used as experimental trials. This dependent measure was analyzed by means of a 2 (anchor: high vs. low) × 2 (time pressure: with vs. without) ANOVA (R-code available online at https://osf.io/6rdpk).

We preregistered multiple outlier and exclusion criteria. First, we included participants' data only if they finished the experiment and provided us with informed consent to analyze their data. Thus, we excluded data from 63 participants. Second, 52 participants who failed to provide their estimates within 7 s in the time pressure condition for at least one of the three experimental trials were excluded. Third, 11 participants who spent more than 60 s with any of the three experimental trials were excluded from the analysis because the longer time indicates that they might have been distracted. Fourth, 26 participants who navigated away from the study page (i.e., tab or window was changed) on any of the three experimental trials were excluded because leaving the page indicates that they might have been using external help (e.g., searching the Internet). Fifth, nine participants who provided no numerical estimates for any of the three experimental trials were excluded. Sixth, for the experimental trials, we classified all responses outside the ranges specified in Table 8 as outliers and excluded 15 participants who provided at least one response that was classified as an outlier from the analysis.

As a manipulation check, we preregistered that we would also analyze participants' mean response times for the three experimental trials with a 2 (anchor: high vs. low) x 2 (time pressure: with vs. without) ANOVA. We predicted a main effect of time pressure with participants who had time pressure responding faster than those without time pressure.

Results

Our final sample consisted of 368 participants (205 female, 162 male, 1 diverse). Their ages ranged from 18 to 75 years, with a median age of 23. This sample size was sufficient to detect effects of $f \geq 0.146$ (or ≥ .021) with 80% power (Faul et al., 2007; see https://osf.io/a7fzb, effect size converted from f to η^2

with effectsize::f_to_eta2 function).

Data Quality Checks

We tested for anchoring effects on an aggregated basis and found a large main effect of anchoring condition on *z*-scores, which were much larger in the high anchor group (M = 0.35, SD = 0.59) than in the low anchor group (M = -0.32, SD = 0.60), $F(1, 364) = 118.42$, $p < .001$, $\eta_p^2 = .25$, 90% CI [.19, .31]. Also for each individual item (*z*-standardized), we found significant anchoring effects: $F(1, 364) = 65.10$, $p < .001$, $\eta_p^2 = .15$, 90% CI [.10, .21] for the third item; $F(1, 364) = 40.74$, $p < .001$, $\eta_p^2 = .10$, 90% CI [.06, .15] for the fourth item; and $F(1, 364) = 38.54$, $p < .001$, $\eta_p^2 = .10$, 90% CI [.05, .15] for the fifth item (see Figure 14).

For *z*-standardized response times, we confirmed that our manipulation resulted in a significantly lower estimation time in the condition with time pressure (M = 4.18 s, SD = 0.80 s) than in the condition without (M = 9.92 s, SD = 6.42 s), $F(1, 364) = 143.42$, $p < .001$, $\eta_p^2 = .28$, 90% CI [.22, .34].

Hypothesis Tests

The interaction between the anchor and time pressure conditions was not significant, $F(1, 364) = 0.53$, $p = .469$, $\eta_p^2 < .01$, 90% CI [.00, .02] (see Figure 15). That is, the anchoring effect was not significantly larger in the condition with time pressure than in the condition without time pressure. For the sake of completeness, the main effect of time pressure was also not significant, $F(1, 364) = 0.03$, $p = .865$, $\eta_p^2 < .01$, 90% CI [.00, .01]. We did not conduct any exploratory analyses beyond the item-by-item ANOVAs reported as data quality checks above.

Discussion

We investigated the influence of time pressure on the anchoring effect for the three classical test items from Jacowitz and Kahneman (1995) listed in Table 8. With self-generated anchors for pleasure indices, Yik et al. (2019) found an asymmetrical effect of time pressure on anchoring. In their third experiment, the anchoring effect increased significantly with time pressure only in the high anchor condition but not in the low anchor condition. Despite the successful manipulation of time pressure, by contrast, we did not find evidence for differences in anchoring effect strength across the time pressure conditions. Accordingly, our results can be taken as evidence against unidirectional serial adjustments from experimenter-provided anchors across time in both conditions.

The origin of example answers plays a crucial, potentially consequential role in terms of paradigmatic construction. For social reasons (e.g., building relationships, reciprocity), we would expect differences in the influence of external numerical input, depending on whether it was generated by experimenters (e.g., an allegedly random number generator; Tversky and Kahneman, 1974) or by peers (Deutsch and Gerard, 1955; Rader et al., 2017). In advice taking studies, researchers typically find that participants egocentrically discount information provided by social others (Yaniv and Kleinberger, 2000). Although we did not ask for independent initial estimates, presenting anchors as responses from previous participants rendered our study conceptually similar to an advice taking experiment (see Bonaccio and Dalal, 2006, for a review). Accordingly, the null effect at hand might also be due to constant amounts of conservatism across the time pressure conditions. In contrast to how much participants adjust away from anchors, time pressure should not affect how egocentrically they behave.

Although a sensitivity power analysis indicated that the study was sufficiently powered, we did not reach our target sample size. According to an a priori power analysis, we were missing 11 participants from the condition with time pressure and high anchors. Additionally, we were missing seven participants from the condition with time pressure and low anchors and seven from the condition without time pressure and with high anchors. Imbalanced data exclusions due to the time pressure manipulation might indicate that 7 s was not enough time to read the question, type an answer, *and* push the continue button. Future versions of our experiment might thus prevent imbalanced data exclusions by replacing the button with a keyboard command for continuing to the next item to facilitate responding.

Deviations From the Preregistration

We did not deviate from the preregistration.

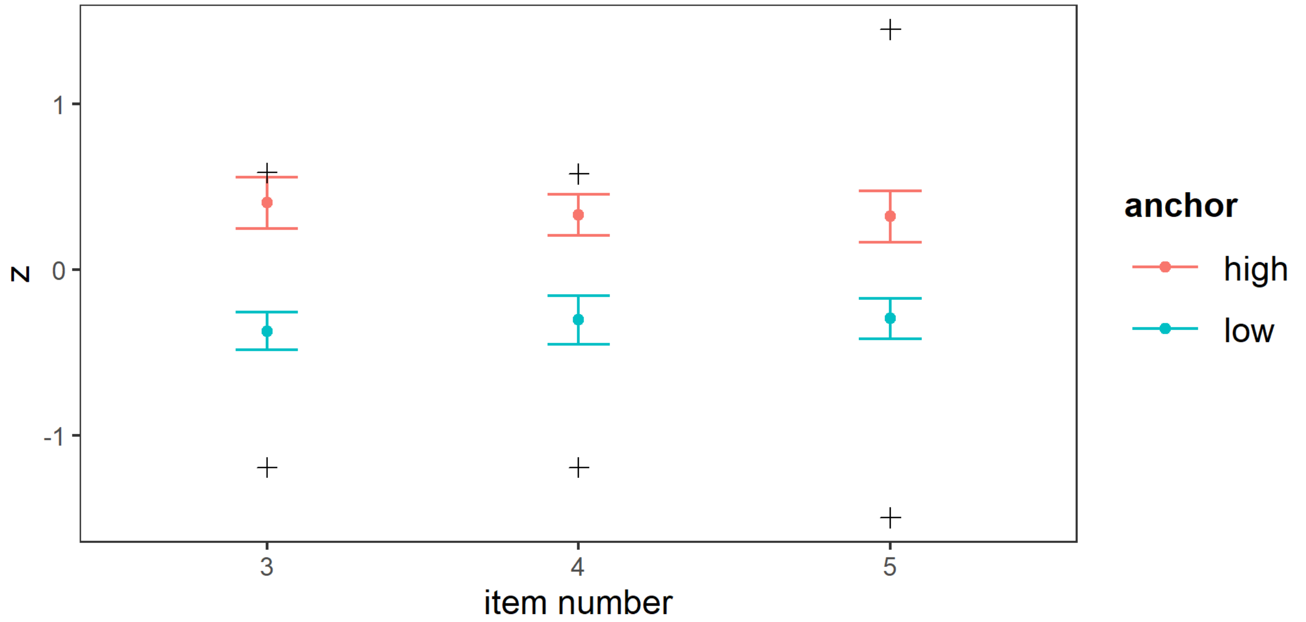
Mini Meta-Analysis

In none of our four studies did we find an effect of time pressure on the strength of the anchoring effect. To test whether this lack of effect was due to a lack of power in the single studies, we aggregated the results and conducted a mini meta-analysis. Note that there is no file-drawer effect in this sample of studies, as the authors declare that they reported all their studies that involved anchoring and a time manipulation. An overview of all the effects is available online (https://osf.io/6sve5).

We converted effect size reports into correlation coefficients (see https://osf.io/znxaf for R-code). For

Figure 14

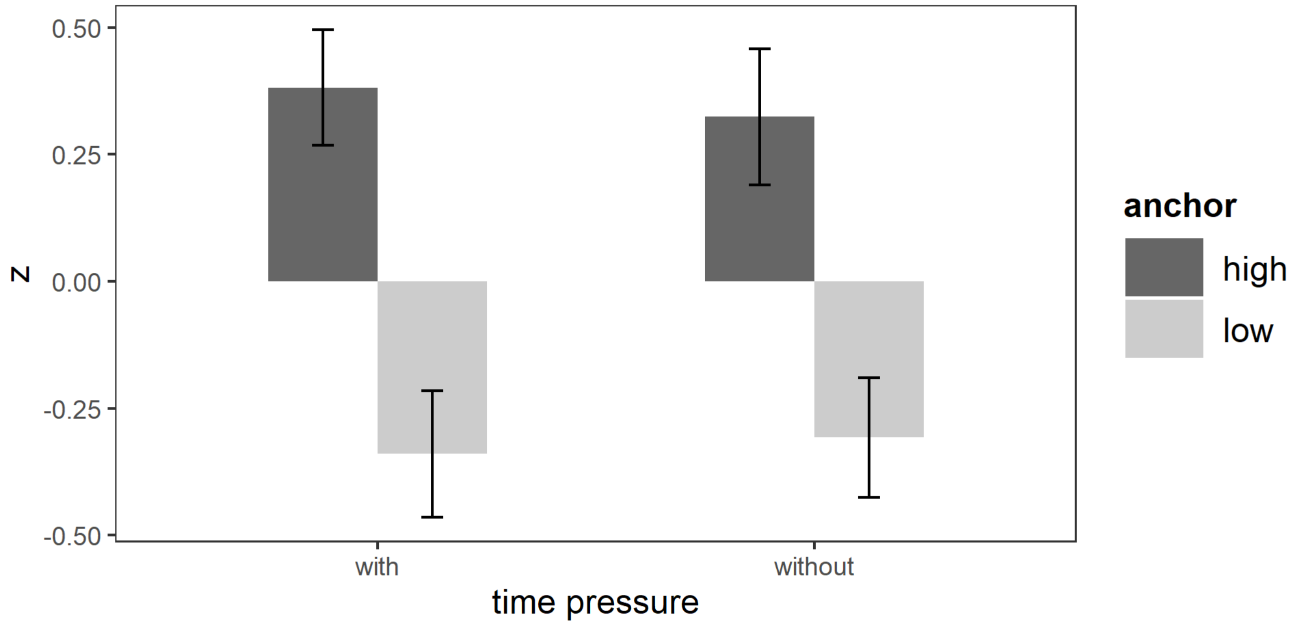
Anchoring Effects for the Three Test Items



Note. For each item, + represents low and high anchors, and dots with error bars represent mean *z*-scores with 95% confidence intervals for the low and high anchor groups.

Figure 15

Mean Susceptibility to Anchoring Across Time Pressure and Anchor Conditions



Note. The bars represent mean *z*-values for the low and high anchor groups with the error bars representing 95% confidence intervals.

Study 2, we used the anchoring effect reported in the exploratory analyses because our hypothesis test using the standardized adjustment scores was not suitable for a meaningful effect size computation. Average effect sizes were computed via a random-effects meta-analysis using the R package metafor version 3.4.0 (Viechtbauer, 2010). Other packages used for the analyses and conversions were openxlsx (Schauberger and Walker, 2021), esc (Lüdecke, 2018a), and psychometric (Fletcher, 2022). Anchoring effects were coded on an item-by-item basis and were thus nested in the study. Time effects were coded on the study level, and no nesting was required. The average anchoring effect size was $r_{anchoring} = 0.191$, $p = .351$, 95% CI [-.211, .594], $N_{total} = 898$. There was large heterogeneity between anchoring effects, $Q(8) = 177.61$, $p < .001$. Time effects were not significantly larger than zero, $r_{time} = 0.02$, $p = .540$, 95% CI [-.045, .086], $N_{total} = 898$, and were homogeneous, $Q(3) = 3.38$, $p = .337$. Blobbograms (forest plots) of the anchoring and time effects are provided in Figure 16 and Figure 17, respectively.

General Discussion

To test the insufficient adjustment model hypothesis that anchoring effects are stronger when there is little time to come up with an estimate, we reported four preregistered studies that varied with respect to the anchoring items, response scales, and time manipulation. Anchoring effects were robust in classical paradigms but when we created a paradigm that is more adequate to test the main hypothesis, the deviations caused anchoring effects to disappear (i.e., responses as 2AFC or on a visual response scale instead of an open numeric field). Time pressure effects on anchoring could not be found in any of the four studies, and the meta-analytic effect size for the time effect was indistinguishable from zero. Anchoring effects are closely tied to the original paradigm. The authors, who come from two independent research groups, therefore strongly believe that the current version of the insufficient adjustment model is wrong. There is no evidence for unidirectional adjustment from anchors over time. Moreover, given our results, past effects of time pressure on the strength of the anchoring effect are inconclusive, imprecise, and unlikely to replicate.

Limitations

The insufficient adjustment model is agnostic about the unit of time, that is, whether adjustment takes a few microseconds or many days. With varying time manipulations, we tried to triangulate the unit of time: For example, in Study 2, we varied the time to range from 4 to 22 s, whereas participants in Study 3 had no time limit or 1 s. Although the vagueness of the insufficient adjustment model still allows it to be rescued from our repeated hypothesis rejections by arguing that adjustment might take only a few nano- or femtoseconds, we do not believe that this is the case. Epley and Gilovich (2001) demonstrated that people possess the interospective ability to report adjustment, and Frech et al. (2020) and Röseler et al. (2020) moreover showed that participants could report each adjustment step when asked to do so (and when asked to proceed in a unidirectional manner). Importantly, an adjustment process that occurs in under 1 s is difficult to investigate using the classical anchoring paradigm.

Given that our hypothesis was an interaction between anchor and time conditions, its effect size could be difficult to detect and the power from our studies insufficient. We applied within-subjects designs to studies 1-3 but did not achieve the target sample size in study 3. To compensate for the lack of power, we ran a mini meta-analysis, which also did not find an overall effect of time. Taken together, we do not expect high-powered studies to find support for the time hypothesis.

Finally, we had to deviate from our preregistrations in some respects. Note that there are very rare cases where no deviations from preregistrations are made (Claesen et al., 2021) and that we discussed each deviation with respect to its reason and effect on the results. None of these deviations affected the results. We invite other researchers to use our materials or data to test further hypotheses about anchoring or time.

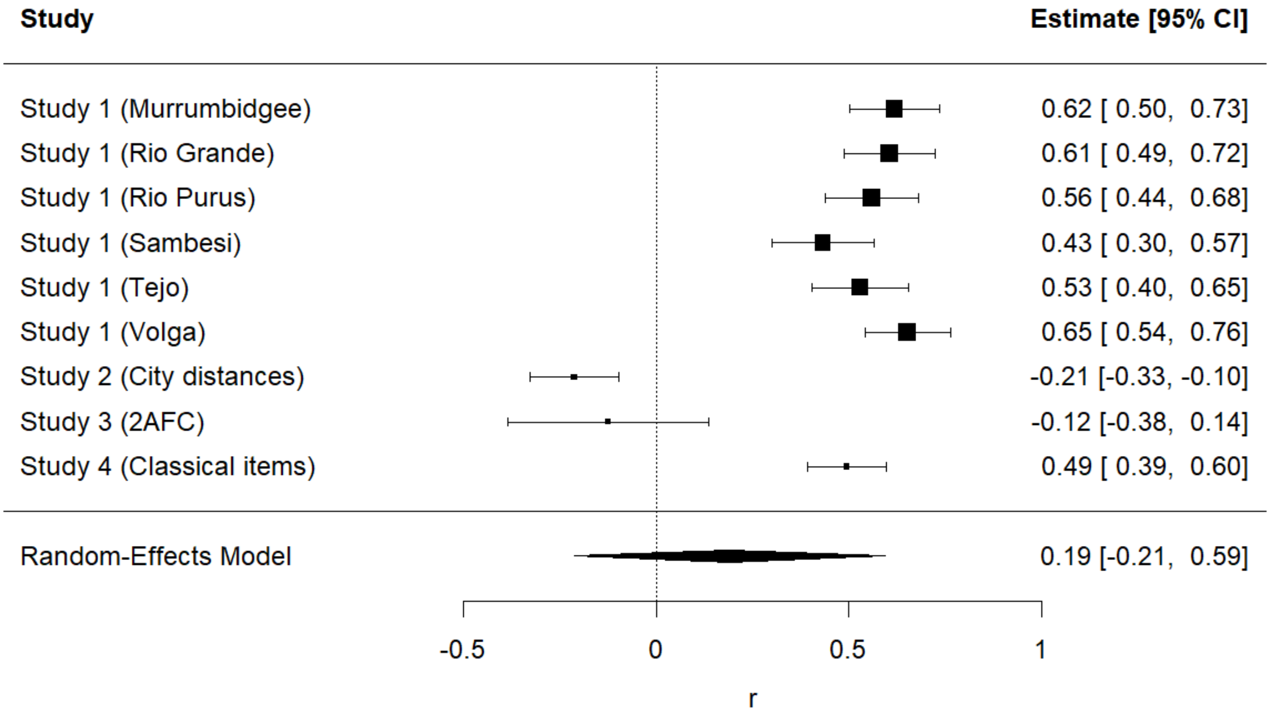
Future Directions

Despite our results that the time hypothesis of anchoring was repeatedly rejected, we still believe some version of the insufficient adjustment model holds the highest potential for anchoring research: It is one of the few models that can be and has been formalized, and it has the highest degree of falsifiability due to its nature as a process. However, we suggest an important alteration to the model in light of our results, namely, to discard the assumption that *adjustments are unidirectional*. We went to great lengths to encourage unidirectional adjustment by framing the anchors as "lower than [anchor]" or "higher than [anchor]" in Studies 1 through 3 (see also Simmons et al., 2010). Still, participants might have adjusted away from the anchor in a first step and back toward the anchor in a subsequent step. Adjustment processes suggested by a *bidirectional adjustment model* are presented in Figure 18.

Incorporating a back-and-forth adjustment into the model would also make the recent replication failures of Epley and Gilovich's findings less dramatic (Röseler, Bögler, et al., 2022). That is, cognitive load or need for

Figure 16

Forest Plot of Anchoring Effect Sizes for All Reported Studies



Note. Larger values of *r* indicate higher estimates for high anchors than for low anchors. Effect sizes were converted to Bravais-Pearson correlation coefficients (see https://osf.io/6sve5 for test statistics and the conversion code).

cognition might not have affected the strength of anchoring because it may have affected only the number of adjustment steps but not the final adjustment distance. Adjustment modeling in 2AFC paradigms could further be conducted by using drift-diffusion models (e.g., Bogacz et al., 2006). Auxiliary hypotheses, such as the equality of step durations or discreteness of steps, deserve closer examination, too. On a similar note, although a frog's aim is to cross the street, it might still jump back a few meters occasionally.

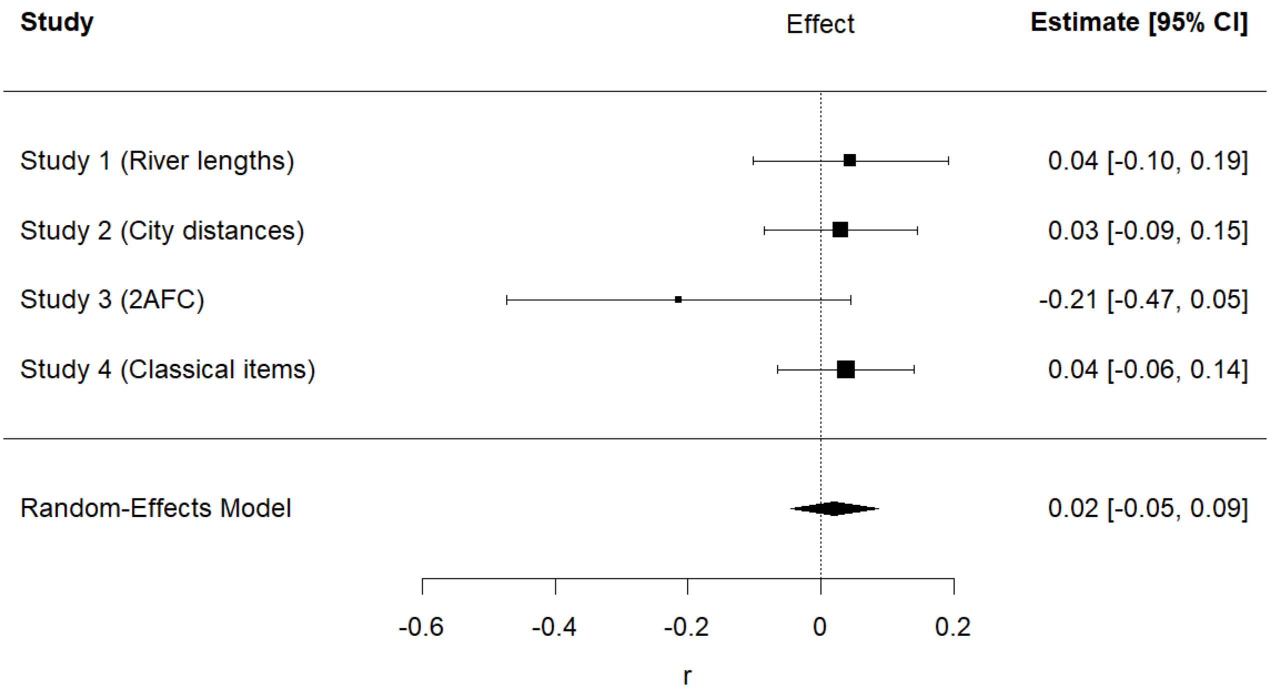
Transparency and Openness Statement

This report is an exhaustive report on all data available from research project relating to the topic, where at least one of the authors was principal investigator, or have otherwise the right to publish the results. This includes not only null findings, or unexpected findings, but also studies that are suspected to have failed, with careful explanation of the circumstances of the failure (e.g., experimental error, failed manipulation check). The context surrounding how these data were collected, and if they are somehow connected to already published studies (e.g., dropped experiments) is carefully explained.

All materials are openly available at https://osf.io/8cwpy/. These include the preregistered and final analysis scripts, the data sets, and the preregistrations. The authors declare that they have emptied their file drawer on anchoring and time pressure studies. That is, all their studies that included multiple subjects, anchoring, and time manipulations are reported in this manuscript. Through personal communications with peers, we know of one additional study that tested the time hypothesis of anchoring using a time pressure manipulation similar to our Studies 1 and 4 and was part of an unpublished Bachelor's thesis. Apart from the fact that time pressure did not affect

Figure 17

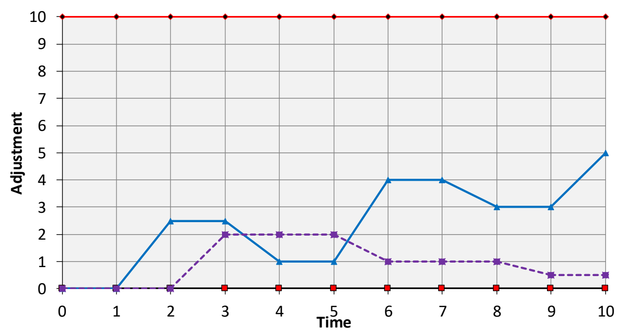
Forest Plot of Time Effect Sizes on Anchoring for All Reported Studies



Note. Effect sizes were converted to Bravais-Pearson correlation coefficients (see https://osf.io/6sve5 for test statistics and the conversion code).

Figure 18

Bidirectional Adjustment Model for Anchoring Paradigms



anchoring effect sizes, no further details about this study have been disclosed to us.

Data from Studies 1, 2, and 4 are included in the Open Anchoring Quest data set (Röseler, Weber, Helgerth, et al., 2022; https://metaanalyses.shinyapps.io/OpAQ).

Author Contact

- Lukas Röseler (corresponding author, lukas.roeseler@uni-muenster.de): 0000-0002-6446-1901
- Lisa Incerti: 0000-0001-6759-0066
- Tobias Rebholz: 0000-0001-5436-0253
- Christian Seida: 0000-0001-8884-2736
- Frank Papenmeier: 0000-0001-5566-9658

We thank Mia Degro for her help in overseeing Study 2 and Jane Zagorski for language editing.

Conflict of Interest and Funding

The authors declare that they have no conflicts of interest. The authors report no financial support for this work.

Author Contributions

Contributions according to CRediT

1. Lukas Röseler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing
2. Lisa Incerti: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing
3. Tobias R. Rebholz: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – review & editing
4. Christian Seida: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – review & editing
5. Frank Papenmeier: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing

Contributions to specific parts of the article

- LR planned and conducted Study 1. He supervised LI and CS in Studies 2 and 3, respectively. LR created the initial draft of the manuscript apart from Studies 3 and 4 and conducted the mini meta-analysis.
- LI planned and conducted Study 3 under the supervision of LR. LI reported Study 3.
- CS planned and conducted Study 2 under the supervision of LR.
- TRR contributed to the study design and reported Study 4.
- FP planned, programmed, and supervised the data collection for Study 4. FP and TRR analyzed Study 4.
- All authors reviewed and edited the final manuscript.

Open Science Practices

This article earned the Preregistration, Open Data, Open Materials, and Open Code badge for preregistering the hypothesis and analysis before data collection, and for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73–106. https://doi.org/10.1162/00335530360535153

Arnold, A. T., & Evans, B. (2021). Bsda: Basic statistics and data analysis (r package version 1.2.1) [computer software].

Bahník, Š. (2021). Anchoring does not activate examples associated with the anchor value. *Advance online publication*. https://doi.org/10.31234/osf.io/4j5wb

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, *107*(1), 66–68. https://doi.org/10.1016/j.econlet.2009.12.028

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects. In *Advances in experimental social psychology* (pp. 319–373, Vol. 42). https://doi.org/10.1016/s0065-2601(10)42006-7

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Brodeur, A., Cook, N. M., Hartley, J. S., & Heyes, A. (2024). Do preregistration and preanalysis plans reduce p-hacking and publication bias? evidence from 15,992 test statistics and suggestions for improvement. *Journal of Political Economy Microeconomics*, *2*(3), 527–561.

Bystranowski, P., Janik, B., Próchnicki, M., & Skórska, P. (2021). Anchoring effect in legal decision-making: A meta-analysis. *Law and Human Behavior*, *45*(1), 1–23. https://doi.org/10.1037/lhb0000438

Champely, S. (2020). Pwr: Basic functions for power analysis. r package version 1.3-0 [computer software].

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

Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, *8*, 211037. https://doi.org/10.1098/rsos.211037

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of Abnormal Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408

Eisenberger, N. I., Lieberman, M. D., & Satpute, A. B. (2005). Personality from a controlled processing perspective: An fmri study of neuroticism, extraversion, and self-consciousness. *Cognitive, Affective Behavioral Neuroscience*, *5*(2), 169–181. https://doi.org/10.3758/CABN.5.2.169

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality Social Psychology Bulletin*, *32*(2), 188–200. https://doi.org/10.1177/0146167205282152

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396. https://doi.org/10.1111/1467-9280.00372

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality Social Psychology Bulletin*, *30*(4), 447–460. https://doi.org/10.1177/0146167203261889

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18*(3), 199–212. https://doi.org/10.1002/bdm.495

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311–318. https://doi.org/10.1111/j.1467-9280.2006.01704.x

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences [http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower3-BRM-Paper.pdf]. *Behavior Research Methods*, *39*(2), 175–191.

Fletcher, T. D. (2022). Psychometric: Applied psychometric theory (r package version 2.3) [computer software].

Fox, J., & Weisberg, S. (n.d.). *An r companion to applied regression (3rd edition)* [https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html].

Frech, M.-L., Loschelder, D. D., & Friese, M. (2019). How and why different forms of expertise moderate anchor precision in price decisions. *Experimental Psychology*, *66*(2), 165–175. https://doi.org/10.1027/1618-3169/a000441

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., Kahneman, D., & Mochon, D. (2010). Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology*, *20*(1), 17–19. https://doi.org/10.1016/j.jcps.2009.12.004

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimen-*

tal Psychology. General, *141*(1), 124–133. https://doi.org/10.1037/a0024006
Gagolewski, M. (2022). Stringi : Fast and portable character string processing in r. *Journal of Statistical Software*, *103*(2). https://doi.org/10.18637/jss.v103.i02
Green, P., & MacLeod, C. J. (2016). Simr : An r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504
Gryz, J., & Krauze, D. (2008). Mortality of vertebrates on a road crossing the biebrza valley (ne poland). *European Journal of Wildlife Research*, *54*(4), 709–714. https://doi.org/10.1007/s10344-008-0200-0
Harrel, F. E. (2020). Hmisc: Harrell miscellaneous. r package version 4.4-0 [computer software].
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 [Advance online publication]. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000644
Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review*, *55*(6), 297–313. https://doi.org/10.1037/h0056721
Højsgaard, S., & Halekoh, U. (2018). Doby: Groupwise statistics, lsmeans, linear contrasts, [computer software].
Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166. https://doi.org/10.1177/01461672952111004
Kassambara, A. (2020). Ggpubr: 'ggplot2' based publication ready plots. r package version 0.4.0 [computer software].
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13
Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z
Lammers, J., & Burgmer, P. (2017). Power increases anchoring effects on judgment. *Social Cognition*, *35*(1), 40–53. https://doi.org/10.1521/soco.2017.35.1.40
Lee, C.-Y., & Morewedge, C. K. (2021). Noise increases anchoring effects [9567976211024254]. *Psychological Science*. https://doi.org/10.1177/09567976211024254
Leiner, D. J. (2019). Sosci survey (version 3.1.06) [computer software].
Li, L., Maniadis, Z., & Sedikides, C. (2021). Anchoring in economics: A meta-analysis of studies on willingness-to-pay and willingness-to-accept. *Journal of Behavioral and Experimental Economics*, *90*, 101629. https://doi.org/10.1016/j.socec.2020.101629
Lieder, F., Griffiths, T. L., M Huys, Q. J., & Goodman, N. D. (2018a). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin Review*, *25*(1), 322–349. https://doi.org/10.3758/s13423-017-1286-8
Lieder, F., Griffiths, T. L., M Huys, Q. J., & Goodman, N. D. (2018b). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin Review*, *25*(2), 775–784. https://doi.org/10.3758/s13423-017-1288-6
Long, J. A. (2020). Jtools: Analysis and presentation of social scientific data. r package version 2.1.0 [computer software].
Lüdecke, D. (2018a). Esc: Effect size computation for meta analysis [Advance online publication]. https://doi.org/10.5281/zenodo.1249218
Lüdecke, D. (2018b). Sjstats: Statistical functions for regression models (version 0.18.2) [computer software].
Miller, G. A., Galanter, E., & Pribram, K. H. (1986). *Plans and the structure of behavior*. Adams-Bannister-Cox.
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., & 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
Navarro, D. J. (2015). *Learning statistics with r: A tutorial for psychology students and other beginners* (Version 0.6). University of New South Wales.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*, 84–97.

Orr, D., & Guthrie, C. (2006). Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio State Journal on Dispute Resolution*, *21*(3), 597–628.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, *11*(8), e12329. https://doi.org/10.1111/spc3.12329

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, *130*, 26–43. https://doi.org/10.1016/j.obhdp.2015.05.004

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

Revelle, W. (2018). Psych: Procedures for personality and psychological research, version = 1.8.3 [computer software].

Röseler, L. (2021). *Anchoring effects: Resolving the contradictions of personality moderator research* [Doctoral dissertation, University of Bamberg] [https://osf.io/34sv6, https://doi.org/10.20378/irb-49951].

Röseler, L., Bögler, H. L., Koßmann, L., Krueger, S., Bickenbach, S., Bühler, R., Della Guardia, J., Köppel, L.-M., Ponader, S., Roßmaier, K., & Sing, J. (2022). Replicating epley and gilovich: Need for cognition, cognitive load, and forewarning do not moderate anchoring effects. *Advance online publication*. https://doi.org/10.31234/osf.io/bgp3m

Röseler, L., & Schütz, A. (2022). Hanging the anchor off a new ship: A meta-analysis of anchoring effects. *Advance online publication*. https://doi.org/10.31234/osf.io/wf2tn

Röseler, L., Schütz, A., Blank, P. A., Dück, M., Fels, S., Kupfer, J., Scheelje, L., & Seida, C. (2021). Evidence against subliminal anchoring: Two close, highly powered, preregistered, and failed replication attempts. *Journal of Experimental Social Psychology*, *92*, 104066. https://doi.org/10.1016/j.jesp.2020.104066

Röseler, L., Schütz, A., Dolling, I. K., Friedinger, K., Hösch, Y., Hügel, J. C., Igna, E., Karg, A., Kühling, S., Schäffer, D., Zhou, A., & Röseler, J. J. (2020). The stepwise anchoring paradigm: Measuring reliable components of anchoring and adjustment as the next step in moderator research. *Advance online publication*. https://doi.org/10.31234/osf.io/hjbwp

Röseler, L., Weber, L., Stich, E., Helgerth, K., Günther, M., Wagner, F.-S., & Schütz, A. (2022). Measurements of susceptibility to anchoring are unreliable: Meta-analytic evidence from more than 50,000 anchored estimates [https://doi.org/10.31234/osf.io/b6t35].

Schauberger, P., & Walker, A. (2021). Openxlsx: Read, write and edit xlsx files (version r package version 4.2.5) [computer software].

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Schultze, T., & Loschelder, D. D. (2020). How numeric advice precision affects advice taking [Advance online publication]. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2211

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/ exclusion model of assimilation and contrast effects in social judgement [http://nbn-resolving.de/urn:nbn:de:0168-ssoar-68883]. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments*. Lawrence Erlbaum Associates.

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., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology*, *99*(6), 917–932. https://doi.org/10.1037/a0021540

Song, A., Faugeras, O., & Veltz, R. (2019). A neural field model for color perception unifying assimilation and contrast. *PLoS Computational Biology*, *15*(6), e1007050. https://doi.org/10.1371/journal.pcbi.1007050

Spicer, J., Zhu, J.-Q., Chater, N., & Sanborn, A. N. (2022). Perceptual and cognitive judgments show both anchoring and repulsion. *Psychological Science*, *33*(9), 1395–1407. https://doi.org/10.1177/09567976221089599

Sweeny, T. D., Grabowecky, M., & Suzuki, S. (2011). Simultaneous shape repulsion and global assimilation in the perception of aspect ratio. *Journal of Vision*, *11*(1), 16. https://doi.org/10.1167/11.1.16

Team, R. C. (2018). R [computer software] [https://www.R-project.org/]. *R Foundation for Statistical Computing*.

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

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (Second edition). Springer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Francois, R., Henry, L., & Müller, K. (2018). Dplyr: A grammar of data manipulation. r package version 0.7.6 [computer software].

Wickham, H., & Girlich, M. (2022). Tidyr: Tidy messy data. r package version 1.2.0 [computer software].

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

Yik, M., Wong, K. F. E., & Zeng, K. J. (2019). Anchoring-and-adjustment during affect inferences. *Frontiers in Psychology*, *9*, 2567. https://doi.org/10.3389/fpsyg.2018.02567

Zhang, Z., & Mai, Y. (2022). Webpower: Basic and advanced statistical power analysis (r package version 0.7) [computer software].

Zong, Y., & Guo, X. (2022). An experimental study on anchoring effect of consumers' price judgment based on consumers' experiencing scenes. *Frontiers in Psychology*, *13*, 794135. https://doi.org/10.3389/fpsyg.2022.794135