Anchoring Effects: Resolving the Contradictions of Personality Moderator Research

▲ Hochschule Harz

Hochschule für angewandte Wissenschaften

vorgelegt von
Lukas Röseler

Hofheim am Taunus, den 16.09.2020

Tag der mündlichen Prüfung: 10. Mai 2021

Dekan: Universitätsprofessor/-in Dr. Jörg Wolstein

Betreuerin: Universitätsprofessorin Dr. Astrid Schütz

Weitere Gutachterin: Hochschulprofessorin Dr. Ulrike Starker

III

Acknowledgements

English Summary

Anchoring effects, that is, the attraction of numerical judgments toward a previously considered number, are among the most robust findings in social psychology. At the same time, explaining the effect and its nuances such as moderator effects has not yet been accomplished despite numerous different models of anchoring. During the past decade, personality moderators of anchoring effects have gained attention and replications of findings beyond the classical anchoring effects have been conducted. I summarize the state of personality moderator research on anchoring effects. I systematizing the eight most thoroughly investigated personality traits (Big Five, cognitive ability, cognitive reflection, and self-control) into three broad paths (exploratory, based on theories of judgment and decision making, and based on anchoring theories). Meta-analyses for each of the potential moderators revealed that there is currently no evidence of any personality moderator of anchoring. The most plausible reason for the absence of personality moderators is that individuals' susceptibility to anchoring cannot be measured reliably. I discuss possible solutions, such as parameters derived from the anchoring and adjustment model.

Keywords: anchoring, personality, moderation, reliability, meta-analysis

German Summary

Ankereffekte (der Einfluss einer in Erwägung gezogenen Zahl auf nachfolgende numerische Schätzungen) gehören zu den robustesten Befunden der Sozialpsychologie. Gleichzeitig ist es bis jetzt trotz vieler Ankermodelle nicht gelungen, das Phänomen mit seinen Nuancen und Moderatoren zu erklären. Während der letzten 13 Jahre lag der Fokus der Forschung zu Ankereffekten auf den Zusammenhang zwischen der Anfälligkeit für Ankereffekte und Persönlichkeitseigenschaften. Ich fasse den Stand dieser Forschung systematisch zusammen, indem ich die acht am stärksten untersuchten Persönlichkeitseigenschaften (Big Five, Intelligenz, kognitive Reflexion, und Selbstkontrolle) in drei Forschungsprogramme (exploratorische Forschung, Forschung basierend auf Modellen des Entscheidens und Urteilen, und Forschung basierend auf Ankermodellen) zusammenfasse. Meta-Analysen für alle acht Moderatoren ergaben, dass keiner von ihnen einen mit der Anfälligkeit für Ankereffekte zusammenhängt. Der dafür plausibelste Grund ist, dass die Anfälligkeit für Ankereffekt nicht reliabel gemessen werden kann. Ich diskutiere verschiedene Lösungen für dieses Reliabilitätsproblem, wie zum Beispiel die theoriebasierte Herleitung neuer Parameter.

Stichwörter: *Ankereffekt, Persönlichkeit, Moderation, Reliabilität, Meta-Analyse*

Preamble

This work is a publication-based dissertation consisting of four articles that are

referenced and summarized in the text and one additional article that is unrelated to anchoring.

The submitted versions of the articles are attached in Appendix II. The submitted or accepted

articles included are as follows:

- Röseler, L., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce judgment adjustment for both internally and externally generated anchors? *Journal of Experimental Social Psychology*, *87*, 103942. https://doi.org/10.1016/j.jesp.2019.103942
- Röseler, L., Schütz, A., Blank, P. A., Dück, M., Fels, S., Kupfer, J., Scheelje, L., & Seida, C. (2020). No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered, and Failed Replication Attempts. *Manuscript submitted for publication*.
- Röseler, L., Schütz, A., & Starker, U. (2019). Cognitive Ability does not and cannot Correlate with Susceptibility to Anchoring Effects. Advance online publication. https://doi.org/10.31234/osf.io/bnsx2
- 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., Ebert, J., Schütz, A., & Baumeister, R., F. (2021). The Upsides and Downsides of High Self-Control: Evidence for Effects of Similarity and Situation Dependency. *Europe's Journal of Psychology*.

This work is not only a summary of my three years of research on anchoring effects but it

also presents new results and integrates findings from the whole field for the first time. I cited the

research articles listed above when appropriate. The introduction is entirely new. Two of the

three pathways to identify potential moderators are new. Pathway 3 is based on ideas by Röseler,

Schütz, Dolling et al. (2020) but is more elaborate. The role of personality sections on the Big

Five and cognitive reflection/impulsiveness are new. The section on cognitive ability

summarizes and integrates the results from Röseler et al. (2019). Solving the reliability issue is a

summary of the findings by Röseler, Schütz, Dolling et al. (2020). The discussion is new as it is

the first time that the large amount of new and old findings are integrated. Appendix I is new.

The other appendices include the research articles listed above.

List of Contents

1. Introduction

Numerical estimates are an integral part of human judgment. Some of the most pressing societal issues essentially boil down to numerical estimates. For example, in how many years should states phase-out fossil-fuels? How many weeks should schools stay closed during a pandemic? How many letters does this sentence have? And what proportion of research articles on anchoring start with a collection of questions? When confronted with the next numerical estimate, consider the random number 42. Among the more certain facts that psychological science has produced is: Your judgment will be drawn toward this number (e.g., Kahneman, 2012).

The phenomenon described above is called *anchoring*, because the number that one has previously considered works like an anchor for one's judgment. Anchoring has been popular in the social sciences since Tversky and Kahneman's (1974) seminal paper on heuristics and biases. Furthermore, books on heuristics and biases such as Kahneman's best selling book "thinking fast and slow" (Kahneman, 2012) made anchoring and many other important psychological findings known to the public. The notion of an anchor as a "context to a target" dates back to the very beginning of empirical psychology (e.g., psychophysics; Fechner, 1876/2013) and has not lost any relevance.

In this work, I summarize the current state and take on challenges of anchoring research. During the last decade, personality moderator research of anchoring effects has been one of the main foci. However, numerous contradictions between theories[1] and findings need to be resolved to evaluate the results of this research programme.

---

[1] The terms model and theory are used interchangeably in this work.

1.1. Variations of Anchoring

Research has yielded numerous paradigms so that talking of "the anchoring effect" (singular) glosses over important differences, such as (1) random anchors (e.g., spinning a fortune wheel; Tversky & Kahneman, 1974) versus informative anchors (e.g., considering a friend's advice; Mussweiler et al., 2000), (2) self-generated anchors (e.g., first thinking of 100°C when being asked for the boiling point of water on Mt. Everest; Epley & Gilovich, 2001), (3) "hidden" anchors (e.g., numbers in the names of products; Critcher & Gilovich, 2008; numbers presented for 17-34 ms; Mussweiler & Englich, 2005; Reitsma-van Rooijen & Daamen, 2006), (4) and sequential anchors (e.g., biasing consecutive judgments by manipulating the order; Frederick & Mochon, 2012). A possible reason for the pluralism of paradigms is that among many psychologists, the view prevails that facts are worth more than theories: "Accumulated evidence for reliable effects is the lasting legacy of science—theories come and go, changing to account for new evidence and making new predictions." (Simons, 2014, p. 79; cf. Muthukrishna & Henrich, 2019). Despite the plurality of methods, anchoring effects are considered reliable and robust.

1.2. Robust and Large Effect Sizes

Anchoring has been called "one of the most reliable and robust results of experimental psychology" (Kahneman, 2012, p. 119) and "one of the most consistent findings in psychology" (Cheek & Norem, 2019, p. 1). As these claims are rarely accompanied by evidence, I report replication results and the effect sizes to support the robustness of anchoring (see Appendix I: The Size of Anchoring Effects).

Not only can anchoring effects easily be replicated, but their effect sizes are very large. To understand the size of anchoring effects, consider common effect size conventions: For

example, Cohen (2013) labelled $d > 0.8$ a large effect based on the distribution of effects from a set of published studies. In a more recent sample of 100 psychological studies that were published in 2008 in one of three important psychology journals and replicated by the Open Science Collaboration (2015), the average replication effect size of all 100 hypothesis tests was $r = .197$ ($SD = 0.257$, p. 5) which corresponds to $d = 0.402$ (small effect). The distribution is right-skewed, that is, large effects are relatively rare. The effect sizes of the replication studies are smaller because there is no publication bias (e.g., Sterling, 1959).

Putting anchoring effects in the context of effect size conventions stresses their effect size. For example, the Many Labs replication studies (Klein et al., 2014) successfully replicated the anchoring effects reported by Jacowitz and Kahnemann (1995) among many other effects. The effect sizes were the largest ones with Cohen's ds between 1 and 2.5 depending on the item (Klein et al., 2014, p. 148). These effect sizes exceed the common effect size conventions by far.[2] Such large effect sizes are rarely found in social psychology. They can be observed for the well-known height difference between men and women ($d = 1.85$), differences in the number of owned pairs of shoes between men and women ($d = 1.07$), or the effect of age (median-split) on the number of years until retirement ($d = 1.49$; for an overview see Simmons et al., 2013, p. 37).

There is one single instance of a failed attempt (Maniadis et al., 2014) to replicate a classical anchoring effect (i.e., the effect of random numbers on willingness to pay; Ariely et al., 2003). The report was heavily criticized for its methodological shortcomings by Simonsohn et al. (2013) because (a) Maniadis et al. (2014) reduced the sample size and thereby statistical power and (b) the replication effect size was not significantly different from the original study's effect

---

[2] To further put the numbers into context, Cohen's ds of 1 and 2.5 correspond to common language effect sizes between approximately 65% and 80%, respectively. In other words, taking only one participant per group, the high anchor participant has higher values than the low anchor participant in up to 80% of all cases. In order to get a significant p-value for such large an effect size with 95% power, only five participants per group are necessary.

size. As a reaction to the failed replication attempt, Simonsohn et al. (2013) applied p-curve

analysis[3] (e.g., Simonsohn et al., 2014a) to the findings of Ariely et al. (2003) and strongly

rejected the hypothesis that anchoring effects are false-positives.

The replicability of anchoring moderators has not received as much attention: Several

moderators have been suggested (e.g., Englich & Soder, 2009; Francis et al., 2018; Lammers &

Burgmer, 2017)[4] but few of them have been replicated. Among the replications of the Open

Science Collaboration (2015) was only one study on an anchoring moderator effect (i.e., on the

effect anchor precision on the strength of anchoring effects) by Janiszewski and Uy (2008). The

results of the replication were extremely similar to the original study. The study was thus

successfully replicated. To sum up, anchoring effects are extremely robust and large. Whether

this holds for moderators of anchoring, however, is uncertain.

1.3. Uncertain Times in Anchoring Research

The replication crisis has hit psychology hard and the confidence in the robustness of

most findings low (e.g., Rohrer et al., 2018; Gordon et al., 2020). Although single estimates such

as replication effect sizes being "half the size of original effect sizes" or "$p < .05$ in 36% of all

studies" do not do justice to the differentiated view of science, it is evident that replications

reveal weaker evidence than original studies in the majority of cases (Open Science

Collaboration, 2015). While anchoring has aged well and the replication crisis has become

evident, some other phenomena that were hardly questioned in before the replication crisis could

---

[3] P-curve analysis is a tool that was developed to detect questionable research practices and correcting meta-analytical effect size estimates for publication bias by Simonsohn et al. (2014a). P-values are uniformly distributed under the H0 and right-skewed under the H1. However, questionable research practices lead to a left-skewed distribution of p-values between 0 and .05 because researchers are incentivized to alter their analyses until their hypothesis is confirmed with p < .05 but not further.

[4] Although the focus in this work is on personality traits as moderators of anchoring, it should be noted that p-curve analyses of moderator effects of power (https://osf.io/68ukn/), mood (https://osf.io/9xkzu/), ego depletion (https://osf.io/qy3bv/), but not precision (https://osf.io/cgxvy/) indicate the absence of evidential value. Replicating the anchoring moderator effects of power, mood, and ego depletion is unlikely.

not be more ambiguous today. In some cases, such as ego depletion, there is not even a consensus anymore about whether the literature provides evidence for the existence of the phenomenon (e.g., Friese et al., 2018). Since controversial and non-replicable findings such as professor priming (O'Donnell et al., 2018, see also Lakens, 2014), religion priming (Gervais et al., 2020) or people being able to predict the future (Bem et al., 2015; LeBel & Peters, 2011), much has changed in the publication of findings in psychology: Standards have changed (e.g., journals recommend or even require preregistration; van 't Veer & Giner-Sorolla, 2016; institutions and journals have signed TOP-guidelines, https://www.cos.io/our-services/top-guidelines), transparency checklists and preregistration templates are being developed (e.g., Aczel et al., 2020), and in some sub disciplines, researchers even suggested a "moratorium of non-preregistered empirical studies" (e.g., Inzlicht & Friese, 2019, p. 374). Pioneers in the field of open and transparent science such as Nelson et al. (2018) even speak of a "renaissance of psychological science".

Can the replication crisis have affected such a robust phenomenon as anchoring? The robustness of anchoring could have led the whole field of anchoring research to seem less prone to problems such as questionable research practices than other fields (e.g., Doyen et al., 2012). While classical anchoring effects have been shown to be robust, a large part of the literature of the past decade consists of moderator effects, such as a negative correlation between susceptibility to anchoring and cognitive ability (Bergman et al., 2010), that have not received as much scrutiny. As I will show, there are non-overlapping theories and numerous contradictions, many of which remain uncommented in the literature. I show that contradictory findings cannot be replicated and emerged due to questionable research practices (Fiedler & Schwarz, 2016; John

et al., 2012), methodical sloppiness (LeBel, 2015), or are simply false-positives. The contradictions to be solved are as follows:

(1) Based on the *Selective Accessibility Model* (the incumbent model of anchoring for the last two decades; e.g., Mussweiler & Strack, 1999a), contradictory hypotheses have been derived that could both be supported: More specifically, it has been argued and demonstrated that ego depletion (i.e., the lack of situational self-control) increases (Banker et al., 2017) but also decreases (Francis et al., 2018) susceptibility to anchoring effects.

(2) Subliminal anchoring, that is, the influence of numbers that were presented for an unnoticeably short time, has been shown to work *only* under time pressure in one study (Reitsma-van Rooijen & Daamen, 2006) versus when people take their time to make an estimate in two other studies (Mussweiler & Englich, 2005).

(3) The state of personality research in anchoring is particularly problematic: A large amount of current research investigates the role of personality traits. Epley and Gilovich (2010) even characterized the current wave of anchoring research as testing moderator variables. The result is a patchwork of directly contradictory findings, for example, three dimensions of the Big Five have been found to correlate with the susceptibility to anchoring positively, negatively, and not at all (for an overview see Cheek & Norem, 2019). Cognitive ability (e.g., SAT-scores or cognitive reflection task scores, Frederick, 2005) seem to be related to anchoring directly (Bergman et al., 2010) but also only indirectly when factors are controlled for (Stanovich & West, 2008).

The main focus of this work is on contradiction (3). In order to assess the state of personality research in anchoring, contradictions (1) and (2) will be resolved.

These conflicts compromise scientific progress: Juxtaposed theories that partly overlap and partly contradict each other (e.g., Mochon & Frederick, 2013, p. 77) and that can hardly be integrated (cf. Simmons et al., 2010) reduce the epistemic connectedness of findings, circumvent critical discourse and thereby reduce systematicity—the very thing that distinguishes science from non-science (Hoyningen-Huene, 2013). The dissolution of the current contradictions would open up an opportunity for scientific progress in the field of anchoring research. For example, a replication of subliminal anchoring is informative regardless the results because it either supports one of the two accounts or none of them. Additionally, resolving contradictions enables inductive theory building (also referred to as "phenomenon-oriented research" or "effects application"; Calder et al., 1981). A theory explaining all current published and contradictory findings on anchoring is unfalsifiable in many domains because it has to make contradictory predictions (e.g., Glöckner & Betsch, 2011; Popper, 1959/2008). The aim of this work is to resolve these conflicts and pave the way for reproducible research on the role of personality in anchoring.

2. Three Pathways to Identify Potential Personality Trait Moderators of Anchoring

I describe the entire research on personality moderators of anchoring by summarizing three different paths to finding personality trait moderators of anchoring effects. To evaluate the evidence, I computed meta-analytical effect size estimates for each of the respective moderators of interest. I provide reasons for potential contradictions. (1) One of the three paths is an exploratory approach and the others derive hypotheses about personality moderators of

anchoring based on (2) judgment and decision making models and (3) anchoring models. An overview of the pathways and the associated models and moderators is provided in Figure 1.

Figure 1

Overview of three personality trait moderator pathways.

Pathway	Model	Moderator
Exploratory		Big Five
Judgment and Decision Making Models	FCIDHBT (Stanovich & West, 2008)	Cognitive Ability
Anchoring Models	Selective Accessibility Model	
	Insufficient Adjustment Model	Cognitive Reflection
		Self-Control
	Scale Distortion Theory	
	Attitude Change Model	
	Incidental Environmental Anchoring	
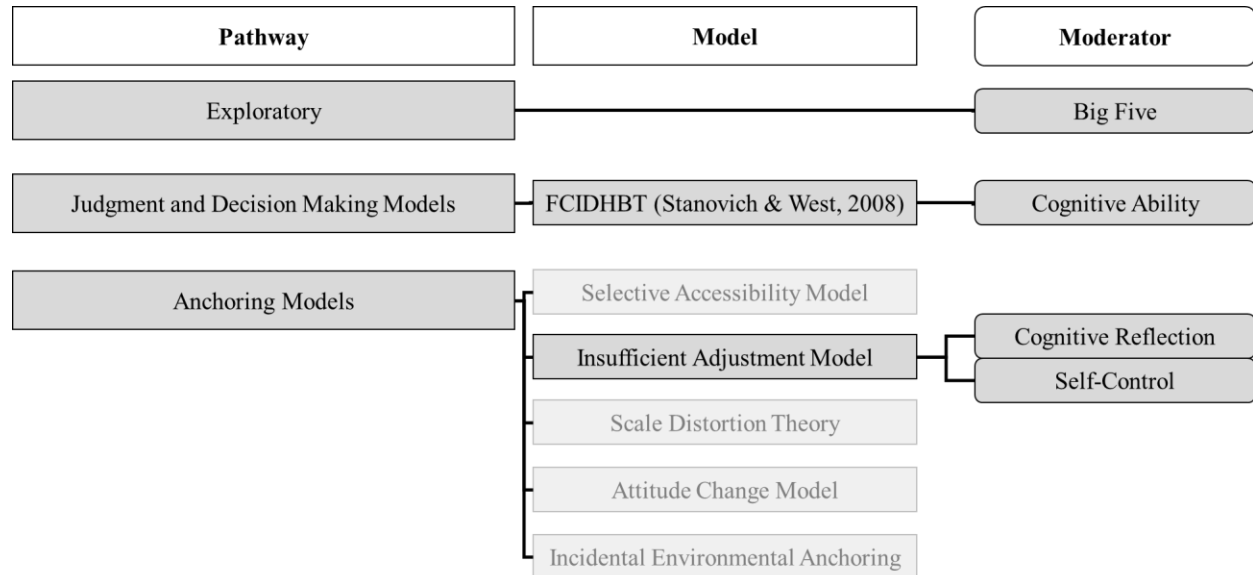
2.1. Exploratory Path

One can deny the usefulness of existing theories or at least evade them in the search for hypotheses on personality traits and simply conduct exploratory research. Followers of this path have thoroughly examined the Big Five (e.g., Furnham et al., 2012; McElroy & Dowd, 2007). Building on a literature review (Cheek & Norem, 2019), I conducted a meta-analysis and funnel-plot analysis to summarize the current state of research and account for potential contradictory findings.

2.2. Judgment and Decision Making Models

The second path makes use of models of judgment and decision making. The only model that can be applied to anchoring and personality was provided by Stanovich and West (2008). Building on their "framework for conceptualizing individual differences on heuristics and biases

tasks" (p. 687), an anchoring task can be designed in which the susceptibility to anchoring effects should be correlated negatively with cognitive ability. I will briefly summarize the framework and the failed efforts undertaken by Röseler et al. (2019) made to test the model with respect to anchoring effects.

2.3. Anchoring Models

The third path makes use of existing anchoring models to derive hypotheses about personality moderators. Deepening the descriptions by Röseler, Schütz, Dolling et al. (2020), I provide an overview of five of the most prominent anchoring models and show that only the anchoring and adjustment model (Epley & Gilovich, 2001) can be used to derive solid predictions about personality moderators of anchoring. Accordingly, cognitive reflection and impulsiveness should be associated with the susceptibility to anchoring.

Originally, anchoring was embedded into research of assimilation and contrast (e.g., Hovland et al., 1957; Sherif et al., 1958) and was interpreted as a measure of communication similar to how scale distortion models anchoring today. It was later hypothesized to generalize to other contexts (Helson, 1964). Since Tversky and Kahneman's (1974) research on heuristics and biases, "insufficient adjustment" had been the incumbent account for the bias. After the first more specific model was introduced by Mussweiler and Strack (1999a, 1999b), Epley and Gilovich (2001) refined the insufficient adjustment account. In the following years, more accounts were suggested (Frederick & Mochon, 2012; Pohl et al., 2003; Wegener et al., 2001) and integrative efforts have been made (e.g., Turner & Schley, 2016). In the following, I discuss the five most prominent theories and provide an overview of them in Table 1. For each of them, I discuss whether and which hypotheses on personality moderators can be derived.

Table 1

Overview of Anchoring Theories

Theory	Content
Anchoring and adjustment	"They 'jump' some amount from the anchor—analogous to a saccade in reading—to a more reasonable value and assess its plausibility. If the new value seems plausible, adjustment stops. If it does not seem plausible, a new jump or saccade is made, the new value is assessed, and so on." (Epley & Gilovich, 2001)
Selective accessibility model	"hypothesis-consistent testing and semantic priming, the model assumes that anchoring effects are mediated by the selectively increased accessibility of anchor-consistent knowledge" (Mussweiler & Strack, 1999a, 1999b)
Scale distortion theory	"The anchor changes how the response scale is used, not how the focal stimulus is perceived. Of importance, we maintain that this holds true even for so-called objective scales (e.g., pounds, calories, meters, etc.). In support of this theory of scale distortion, we show that prior exposure to a numeric standard changes respondents' use of that specific response scale but does not generalize to conceptually affiliated judgments rendered on similar scales." (Frederick & Mochon, 2012)
Attitude change	"Anchors could directly serve as a cue or indirectly influence the information processing that bias judgments toward the anchors. Low-elaboration anchoring is resulted especially during non-thoughtful processes; anchors are treated as a 'hint' to a reasonable answer. High-elaborative anchoring involves thorough thinking that engages judges in more effortful information processing with existing knowledge and hence activate the anchor-consistent information that bias judgments." (Furnham & Boo, 2011)
Numeric priming / incidental environmental anchoring	"Anchor values that are incidentally present in the environment can affect a person's numerical estimates .… these effects were not qualified by participants' expertise in the relevant domain (study 1) or by their ability to subsequently recall the anchor value (study 3)." (Critcher & Gilovich, 2008)

2.3.1. Anchoring and adjustment

Anchoring and adjustment (sometimes also referred to as the "insufficient adjustment model") originally consisted of Tversky and Kahneman's (1974) statements that "people make estimates by starting from an initial value that is adjusted to yield the answer" and that "adjustments are typically insufficient" (p. 1128). Evidence that adjustment actually happens was provided much later by Epley and Gilovich (2001) who found that people verbally reported that they adjusted for self-generated anchors. For example, they interpreted the finding that shaking one's head (versus nodding) increases adjustment (Epley & Gilovich, 2001, Study 2) as support for the adjustment hypothesis. Note, however, that this finding might very well support

numerous other accounts of anchoring. For example, if people have a notion of the true value on the one hand and the anchor on the other hand, an anchor-weighting-model could predict, that shaking one's head decreases weights for the anchor and thus decreases the strength of anchoring although no form of adjustment is assumed to happen. More evidence for adjustment was very recently provided in a study where people were asked to report their adjustment steps (Frech et al., 2020). If the model was wrong, people might have trouble adhering to this request. However, adjustment steps were reported (Frech et al., 2020). When individuals had to estimate the prices of three items, the median number of adjustment steps were three steps for the car and secretaire price estimates and two steps for the toy price estimate. ($N = 146$, see https://osf.io/dq48j/ for the dataset for Study 2 provided by Frech et al., 2020 and https://osf.io/ck4y3/ for the analysis script). The proportion of participants who made only one adjustment step was relatively low (i.e., less than five percent of all participants). Note that in line with Simmons et al. (2010) this also supports that people do not only adjust from self-generated anchors but also from experimenter-provided anchors.

Adjustment alone does not create anchoring. For assimilation effects to emerge, adjustment needs to be insufficient. Further reasons for why adjustment is insufficient were not provided until recently: Decision making under uncertainty probably leads to aversion of extreme decisions (Lewis et al., 2019). Sufficient adjustment from an anchor is an extreme decision, which is why adjustment is usually insufficient. Note, however, that Lewis et al. (2019) did not actually manipulate extremeness aversion.

The insufficient adjustment model is superior to the other models in that it can be easily formalized. Using the cybernetic TOTE-loops (i.e., test, operate, test, exit), it can be proposed that people (1) *test*, whether the anchor is plausible, if it is not (2) *operate*, that is, adjust away

from it and (3) *test* again, and if the adjusted value *is* plausible (4) *exit* the adjustment loop and take the adjusted value as the estimate (see Figure 2). For example, the total adjustment can be expressed as the product of adjustment steps and the mean adjustment step width. There are numerous ways to refine this model. For example, the step size is probably not constant but decreases over time (Frech et al., 2020, see https://osf.io/dq48j/ for data and https://osf.io/ck4y3/ for the analysis code), further variables such as adjustment speed can be added. A formalized model allows for more precise hypotheses (Glöckner & Betsch, 2011) and provides a parsimonious but specific alternative to exploratory studies on the influence of the Big Five on anchoring or complex verbal models with multiple vague parameters (Stanovich & West, 2008).

Recent research on the precision of anchors has made use of the model. Besides precision being an indicator of expertise (Frech et al., 2019; Loschelder et al., 2014; Loschelder et al., 2016; Mason et al., 2013), it could also influence peoples representation of numbers to be more fine-grained (Janiszewski & Uy, 2008). Janiszewski and Uy (2008) argued for fine-grained adjustment by showing that precise anchors lead to precise estimate, which could also be due to some sort of linguistic conventions (Grice, 1982). This alternative explanation was ruled out by Frech et al. (2020), who asked participants in two preregistered studies to indicate all adjustment steps and found that precise anchors led to smaller steps.

Based on this model, Röseler, Schütz, Dolling et al. (2020) derived the hypotheses that impulsiveness should be associated with strong anchoring effects and cognitive reflection should be associated with weak anchoring effects. They argue that high impulsiveness and low cognitive reflection represent the tendency to take for an answer what "springs to one's mind" first. If anchoring is due to individuals exiting the TOTE-loop too early because they accept the

insufficiently adjusted number too readily, then high cognitive reflection and low impulsiveness should be a safeguard for high susceptibility to anchoring.

Figure 2

The insufficient adjustment model of anchoring expressed as a TOTE model.

```
┌──────────────────────┐      No      ┌──────────────────────┐
│ Test whether the value │ ───────────► │ Operate by adding or │
│  is a plausible estimate │ ◄─────────── │  subtracting to the  │
└──────────────────────┘              │        value         │
         │                            └──────────────────────┘
         │ Yes        ╭──────────────╮
         ▼            │  Add 1 to the │
┌──────────────────┐  │  number of   │
│ Exit the adjustment │ │  adjustment  │
│     process       │  │    steps     │
└──────────────────┘  ╰──────────────╯
```
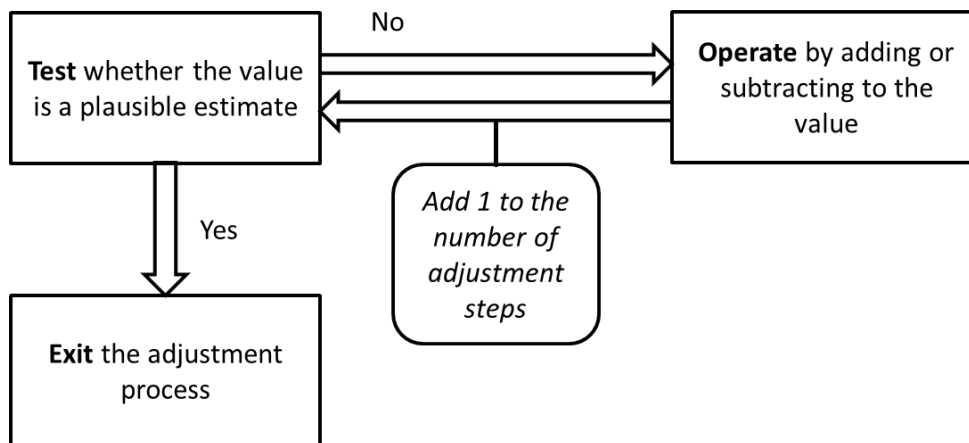
To sum up, although the anchoring and adjustment model has been the first account of anchoring, it has not been refined until recently: Formalization and the development of new paradigms based on TOTE-loops foster the model's practicability.

2.3.2. Selective Accessibility Model

As discussed by Röseler, Schütz, Dolling et al. (2020), priming has been found to be moderated by personality traits such as need for cognition (Petty et al., 2008) or associative thinking (Wood et al., 2011). However, before testing this hypothesis, a closer look at the model reveals several weaknesses.

While priming had been suggested to be a cause of anchoring relatively early (e.g., Kahneman & Knetsch, 1993), reasons for this model to apply were provided by Mussweiler and Strack (e.g., 1999b). They assumed that answering the comparative question (e.g., is the mean temperature in Germany lower or higher than 20°C?) requires considering the value plausible.

This is in line with Spinoza's model of information encoding having a believe stage and a subsequent evaluative stage (and inconsistent with Descartes account that information is first labelled as neither true nor false; Gilbert et al., 1990). In other words, Spinoza's account suggests that in order to *understand* a claim, it needs to be encoded as true at first and then negated in a second step. According to Mussweiler and Strack (e.g., 2000a), this "hypothesis-consistent testing" strategy, should activate information consistent with the anchor (semantic priming). As outlined above, the selective accessibility model has contradicted both the insufficient adjustment model (Schweickart et al., 2020) and scale distortion theory (Mochon & Frederick, 2013) and results could not support the selective accessibility model in either of the cases.

Apart from that, there are even more problematic cases. First, thirty years after the research by Gilbert et al. (1990), it is still disputed whether information is encoded as true at first. For example, Spinoza's account suggests the truth bias should persist even if people have a third option indicating that they are uncertain. However, research findings are consistent with the Descartes' account because there is no bias toward believing false information when uncertainty can be indicated (Street & Kingstone, 2017, Street & Richardson, 2015). There is also evidence that suspending belief is possible (Hasson et al., 2005). The bias towards believing rather than disbelieving early in the process is construed as a best guess by proponents of the Cartesian model (Street & Kingstone, 2017, p. 11).

Second, according to Mussweiler and Strack, the "most compelling evidence stems from an experiment which combined the anchoring paradigm with a lexical decision task" (1999b, p. 140). In two studies they found reaction times to lexical decision tasks for anchor consistent words to decrease in contrast to anchor inconsistent words (Mussweiler & Strack, 2000b, Studies 1 and 2). Taking a closer look, one can easily see that p-values for both results were very close to

5% and sample sizes very similar, which raises concerns about questionable research practices[5] (e.g., John et al., 2012). More specifically, a p-value close to 5% in Study 1 should have been an indicator to increase sample size for further studies (post hoc power for p-values around 5% is around 50%, e.g., Hoenig & Heisey, 2001) but the sample size in Study 2 was almost the same as in Study 1 ($N_{Study\ 1} = 28$, $N_{Study\ 2} = 30$). Much more effective than p-value analysis is assessing the robustness of these findings by direct replication (e.g., Brandt et al., 2014; Hüffmeier et al., 2016). In a nutshell, in a series of four direct and one conceptual replication attempts by Harris et al. (2019), the "most compelling evidence" could not be replicated, and "supportive evidence for selective accessibility in the past using this method is not a strong part of the evidence base for that theory" (p. 41). This poses serious problems for the selective accessibility model.

Among the basic requirements for every scientific model is that it has empirical content. More specifically, it must divide possible statements about empirical facts into (a) one class of statements that are *consistent* with the model and (b) another class of statements that are *inconsistent* with the model (e.g., Popper, 1959/2008, p. 95). However, this is not the case in the domain of anchoring and ego depletion: Ego depletion, which is the decline in state self-control and which is said to resemble a general resource that is needed for all kinds of activities (e.g., holding one's arm in cold water, making decisions, suppressing emotions, Hagger et al., 2010), has been predicted to increase but also decrease the strength of anchoring. Note that recent discussions and failed replication attempts in the field of ego depletion caused a great stir and it is even unclear whether the phenomenon actually exists (Baumeister & Vohs, 2016; Friese et al., 2018; Garrison et al., 2019; Hagger et al., 2016; Inzlicht & Friese, 2019). In relationship to

---

[5]The exact results were $F(1, 26) = 4.53$, $p = .04295$ and $F(1, 28) = 6.57$, $p = .01603$. From having a p-value close to 5% in one study follows that subsequent studies need a larger sample size if they aim at adequate statistical power (i.e., $1-\beta > .5$).

anchoring, it has been argued and shown that "ego depletion weakens central executive control, thereby increasing susceptibility to salient cues" (Banker et al., 2017, p. 2) and should thus increase anchoring. At the same time and agnostic of contradictory findings Francis et al. (2018) argued (and showed) that "when someone spends more time thinking about the anchor value, the anchor value has more and more influence on one's final estimate" and "weakening deliberative processing seems to decrease the anchoring effect" (p. 277). In a high-powered registered report, Röseler, Schütz, Baumeister, and Starker (2020) tried to resolve this contradiction. Applying the TOTE-model to ego depletion led to predictions in line with those of Banker et al. (2017). What is problematic about their argument is that Banker et al. (2017) used the dictator game, which involves much more than classical anchoring paradigms: People had to reallocate money that was assigned by the experimenter to themselves and another person. To prohibit alternative explanations to work, they used classical anchoring tasks that we took directly from the literature on anchoring. Although ego depletion was successfully manipulated and anchoring occurred, no effect of ego depletion on the strength of anchoring could be found. The submitted version of the registered report by Röseler, Schütz, Baumeister, and Starker (2020) can be found in Appendix II (Article 1: Anchoring and Ego Depletion).

An explanation of the contradiction discussed by Röseler, Schütz, Baumeister, and Starker (2020, p. 8) is the use of the dictator game instead of classical anchoring items by Banker et al. (2017) and its involvement of social mechanisms, which have been shown to matter for anchoring effects in negotiations (e.g., Ames & Mason, 2015). Röseler, Schütz, Baumeister, and Starker (2020) dismiss the possibility that questionable research practices might have led to the contradictory findings (p. 3). However, their p-curve analysis featured all p-values reported by Banker et al. (2017), including manipulation checks and robust anchoring effects, which

concealed a potentially skewed p-curve (Bishop & Thompson, 2016). In order to test the concern of questionable research practices more rigorously, I extracted the hypothesis tests (i.e., effects of ego depletion on anchoring) from the reports by Banker et al. (2017) and Francis et al. (2018) and conducted a p-curve analysis. P-curve analysis uses the facts that $p < .05$ is a criterion for publication and that the distribution of p-values from significance tests is uniform given the null hypothesis and right skewed given the alternative hypothesis. If questionable research practices have been used, the distribution should be left skewed (e.g., Simonsohn et al., 2014a, 2014b; Simonsohn et al., 2015). Results from the full and half p-curve do not indicate evidential value. There is no right skew, $Z_{full\ p\text{-}curve} = -1.24$, $p = .108$. Post hoc power for the reported results is very low, $1\text{-}\beta = .31$, 95% CI = [.05, .82]. The full results and the code necessary to reproduce them are available online (https://osf.io/jmgsc/). This suggests that the original findings were indeed false-positives in the first place and effects emerged due to questionable research practices. That explains why the results are contradictory, and why Röseler, Schütz, Baumeister, and Starker (2020) could not find an effect.

To sum up, the selective accessibility model has shaky theoretical foundations, its key finding is not replicable, and it allows for contradictory predictions in the domain of ego depletion. Using it to derive hypotheses on personality moderators is thus risky.

2.3.3. Attitude Change

The important role of attitude has been demonstrated in the light of extreme anchors. As with anchors in the classical assimilation and contrast research, there are different possible effects of extreme anchors (for an overview see also Chapman & Johnson, 1994, p. 225): (1) According to selective accessibility and the insufficient adjustment, extreme anchors should lead to stronger anchoring effects. Extreme anchors should be more likely to increase the accessibility

of information that is consistent with the anchor (selective accessibility) and they require more adjustment (insufficient adjustment). There is no reason for why the anchor strength should increase or decrease with larger distances of the anchor from the true value, which is why selective accessibility and insufficient adjustment suggest no special relationship or a *linear* relationship between anchor extremeness and anchor effect strength. (2) If anchoring operates via the influence on attitudes, extreme anchors should seem less plausible and thus have a weaker effect: "… the outcomes of persuasion processes might look considerably more like 'disconfirmatory search' rather than 'confirmatory search.' Such 'disconfirmation' might often occur when the social perceiver has some reason to 'disagree' with the value suggested by an anchor" (Wegener et al., 2001, p. 67). The relationship would thus be *decreasing*. (3) A third possibility of the relationship between anchor extremity and anchoring strength is *contrast*, that is, extreme high anchors could annihilate anchoring or even lead lower estimates for high anchors. Contrast would be predicted by Gestalt psychological accounts (i.e., if the anchor and the true value are so far apart that they are perceived as two different wholes; King, 1988) and social judgment theories (Hovland et al., 1957; Schwarz & Bless, 1992).

Research found evidence for the decreasing strength of anchoring (Chapman & Johnson, 1994; Thorsteinson, 2011; Wegener et al., 2001), which is why attitude change is deemed a valid model of anchoring. In order to test this model and the role of attitudes for anchoring, I suggest investigating moderators that decrease the strength of anchoring. That is, some people might discount extreme anchors more readily than others. As anchors seem to include a social component (at least in questions such as salary negotiation used by Thorsteinson, 2011), traits such as narcissism (e.g., Schütz et al., 2004) might be important because they have been shown to be associated with discounts of social disapproval (Leder et al., 2020; see also Smeesters et al.,

2009 for the association of social value orientation and priming). In other words, immediate and strong rejection of the negotiation partner's first offer (i.e., an anchor) in a salary negotiation may not be socially desirable. Whereas usually individuals might try to meet the offer halfway, individuals high in narcissism might not and thus be less susceptible to extreme anchors in negotiation situations.

Models systematically integrating the source of the anchor and thus the amount of information or its credibility suggest that elaboration of anchors is consistent with selective accessibility and no elaboration is consistent with numeric priming models (Turner & Schley, 2016, p. 4-5). To sum up, the attitude change view of anchoring is needed to explain effects of the source of the anchor and can be used to predict personality moderator effects *in domains that include social facets such as politeness*. Derivable moderator hypotheses are not valid for classical anchoring effects, such as quantity estimates.

2.3.4. Incidental Anchoring and Subliminal Priming

Most anchoring models assume anchors will only bias judgment if subjects consider the anchor value (e.g., by answering the comparative question). In incidental anchoring, which is also referred to as subliminal anchoring (subliminal meaning in this case below the *attention* threshold), environmental anchoring, numeric priming, or basic anchoring (Wilson et al., 1996), no conscious consideration is necessary. Instead, any number that is perceived biases judgment. For example, the jersey number of an athlete was found to bias judgments about his performance: Participants "who saw Stan Fischer wearing jersey number 94 [vs. number 54] estimated that his probability of sacking the quarterback in his next game was higher" (Critcher & Gilovich, 2008, p. 244). Note that incidental anchoring is not actually an account of anchoring because no theoretical considerations have been provided prior to or developed since the finding

has been published. Instead, it is an exploration of the limits of anchoring (Critcher & Gilovich, 2008, p. 243).

There is another form of subliminal anchoring with even less requirements. In this sort of subliminal anchoring, subliminal refers to below the *perception* threshold (as opposed to below the attention threshold). That is, people are not only unaware of the anchor but they could not even tell what it is if they knew it would be presented during the experiment. Loosely based on subliminal priming and selective accessibility, Mussweiler and Englich (2005) and Reitsma-van Rooijen and Daamen (2006) showed that subliminally presenting an anchor (i.e., for 17 to 34 ms) influences subsequent numeric estimates. Besides being the only two studies on subliminal anchoring, their findings are contradictory with respect to the role of time-pressure: Mussweiler and Englich (2005) instructed participants to take their time to make an estimate and found an effect whereas Reitsma-van Rooijen and Daamen (2006) manipulated time-pressure and found subliminal anchoring to occur under time-pressure, but no when participants took their time.

Replication studies have recently questioned the credibility of these astonishing findings: In incidental environmental anchoring, one of the studies from Critcher and Gilovich (2008) has been unsuccessfully replicated in the Many Labs 2 project (Klein et al., 2018) and a systematic failed attempt at replicating incidental anchors was recently reported by Shanks et al. (2020). Shanks et al. (2020) summarize the evidence for incidental anchoring and concluded that positive findings (e.g., Dogerlioglu-Demir & Koçaş, 2015) are false-positives and probably due to questionable research practices. In subliminal anchoring, we replicated both studies and found no effect at all (Röseler, Schütz, Blank et al., 2020). A preprint of Röseler, Schütz, Blank et al. (2020) is available in Appendix II (Article 2: Subliminal Anchoring).

To sum up, there is currently no convincing evidence for incidental anchoring. Despite numerous attempts, the findings by Critcher and Gilovich (2008) could not be replicated. As there are neither theory nor evidence, incidental anchoring provides no basis for moderator predictions. The past findings were most likely false-positives: P-curve analysis revealed that there is no evidential value[6]. All three findings were $.02 < p < .05$. The power estimate was the lowest possible value, that is, 5% (95% CI = [5%, 50%]).

2.3.5. Scale Distortion

Anchoring is consistently assimilative and thereby living up to its name. Ironically, Frederick and Mochon (2012) propose that at the core of anchoring, contrast effects are happening. More specifically, by considering a large anchor, previously plausible numbers suddenly seem smaller than they are: "1,000 pounds might seem like a reasonable estimate of a giraffe's weight in the absence of an anchor, the consideration of a small anchor (e.g., 15 pounds) may make 1,000 pounds seem unreasonably large" (Lewis et al., 2019, p. 159). Without explaining this contrast effect any further, two conditions of scale distortion have been specified in a refinement of the theory: (1) the anchor and the target must be assessed on a common scale, (2) respondents must compare them (Mochon & Frederick, 2013, see also Röseler, Schütz, Dolling et al., 2020). The model thus sticks out with its relatively specific conditions for anchoring. However, there is no overlap with traits and it can thus not be used for personality moderator search. In other words, scale distortion theory would be consistent with there being no personality moderators.

---

[6] Values used for the analysis were as follows: $t(33) = 2.13$, $t(39) = 2.0249$ (0.0049 were added to the t-value for $p < .05$) as reported by Mussweiler and Englich (2005), and $F(1, 58) = 5.06$ as reported by Reitsma-van Rooijen and Daamen (2006).

2.3.6. Contradictions between Theories

Taken together, the discussed models can explain many findings in the anchoring literature. However, this pluralistic approach bears downsides (see also Turner & Schley, 2016): Ideally, the models can either be integrated or they contradict each other in at least one aspect so that one can decide between models. Instead, each model focuses on a subset of findings which it can explain and ignores phenomena that it cannot predict (e.g., Mochon & Frederick, 2013, p. 77). Only recently have independent researchers (i.e., researchers who did not propose a theory themselves) begun to identify contradictory hypotheses between models.

For example, Schweickart et al. (2020) showed that anchors do not influence judgments when they are evaluated as "bad estimates". This contradicts the selective accessibility model (which assumes an automatic process that cannot be influenced) but it is consistent with the insufficient adjustment model (which can be characterized as a deliberate heuristic).

Frederick and Mochon (2012) presented another case, in which an anchor describing the width of the Brandenburg gate should influence the height estimate according to the selective accessibility model (Strack & Mussweiler, 1997) but not according to scale distortion (Mochon & Frederick, 2013, Study 1; see also Wong & Kwong, 2000).

Finally, Sailors and Heyman (2019) showed that there is a positive correlation between anchoring effect strength and semantic similarity between the target and the anchor, which is consistent with selective accessibility but not with scale distortion. An overview of findings that are consistent with only one of the models (i.e., exclusive consistencies) and findings that are inconsistent with the models is provided in Table 2.

Table 2

Findings that are consistent with only one anchoring model (exclusive consistencies) and

inconsistencies for the five most prominent anchoring models

Model	Exclusive Consistencies	Inconsistencies
Insufficient Adjustment	People make adjustment steps (Epley & Gilovich, 2001; Frech et al., 2020)	Rewarding accuracy does not work well (e.g., Simmons et al., 2010)
Selective Accessibility	Anchoring leads to consistent information being recognized faster (Mussweiler & Strack, 1999b); semantic similarity is associated with anchoring strength (Sailors & Heyman, 2019)s	Evaluative judgment eliminates anchoring (Schweickart et al., 2020); Anchoring is scale specific (Frederick & Mochon, 2012); hypothesis-consistent testing is implausible (Street & Kingstone, 2017); Different effects despite same semantics (Wong & Kwong, 2000); failures to replicate key result of selective accessibility (Harris et al., 2019)
Scale Distortion Theory	Anchoring is scale specific (Frederick & Mochon, 2012)	Self-generated anchors (e.g.,Epley & Gilovich, 2001)
Attitude Change	Anchoring is weaker for extreme anchors (e.g.,Chapman & Johnson, 1994)	Random anchors work, too (Tversky & Kahneman, 1974)
Incidental Environmental Anchoring	Numeric Priming (Critcher & Gilovich, 2008); Subliminal Anchoring (Mussweiler & Englich, 2005)	Contradictory Findings (Reitsma-van Rooijen & Daamen, 2006); original results could not be replicated and are likely due to QRP (Röseler, Schütz, Blank et al., 2020; Shanks et al., 2020)

I advise authors to raise awareness of the many theories and to test new hypothesis on

anchoring across multiple models instead of picking only one model with which the hypothesis is

consistent (e.g., Banker et al., 2017). A reason for the contradictions is that most theories are

mainly investigated by the researchers who proposed those theories. For example, the only

research on scale distortion was conducted by its creators Frederick and Mochon (Frederick &

Mochon, 2012; Mochon & Frederick, 2013, cf. Bahník et al., 2018), most of the research on the

insufficient adjustment model was conducted by Epley and Gilovich (Epley & Gilovich, 2001,

2004, 2005, 2006, 2010), and most of the research on the selective accessibility model was

conducted by Mussweiler and Strack (Mussweiler et al., 2000; Mussweiler & Strack, 1999a,

1999b, 2000a, 2000b, 2000b, 2001a, 2001b; Strack et al., 2016; Strack & Mussweiler, 1997).

3. The Role of Personality for the Susceptibility to Anchoring

Current anchoring research is mainly concerned with moderator detection and the results could hardly be more contradictory. I summarize the state of the research for each of the three paths outlined above (using Exploratory, Judgment and Decision Making Models, and Anchoring Models to predict personality trait moderators of anchoring) and critically evaluate the evidence. To do so, I list a plethora of findings and conduct meta-analytical tests for publication bias and for questionable research practices for eight personality traits (i.e., Big Five, cognitive ability, cognitive reflection, and self-control / impulsivity).

3.1. Meta-Analyses for Path 1: Big Five

The majority of anchoring moderator research has chosen the exploratory path and examined the association between the Big Five (e.g., Boyle et al., 2008, p. 179) and people's susceptibility to anchoring (e.g., Cheek & Norem, 2019, p. 2)[7].

3.1.1. Theory

The Big Five are based on the lexical hypothesis, that is, every important trait of individuals is represented in the language (Galton, 1884). The condensation of descriptive, non-evaluative word from dictionaries through factor analytical methods has yielded the five personality factors neuroticism, extraversion, openness, agreeableness, and conscientiousness across different cultures and times (Allport & Odbert, 1936). Their breadth are an optimal opportunity to reveal correlations with hitherto unknown constructs such as the susceptibility to anchoring and point future research to more specific traits.

---

[7]Note that Cheek and Norem (2019) and Perry and Sibley (2013) did actually derive indirect hypotheses that could not be confirmed. They recommend deriving more specific hypotheses and cite Lammers and Burgmer (2017) as a model example. Unfortunately, a p-curve analysis of the findings by Lammers and Burgmer, revealed that their results are highly unlikely to replicate: The analysis did not indicate evidential value, estimated power was 5%, 95% CI = (5%, 45%). The values used in the p-curve analysis are available online (https://osf.io/68ukn/).

Cheek and Norem (2019) summarized the existing Big Five Anchoring studies and their contradictions: Each of the traits has been found to be unrelated to anchoring in some studies, four have been found to be positively related to anchoring in other studies, and in further studies, three have been found to be negatively related to anchoring (p. 2). In a large preregistered study they conducted themselves, Cheek and Norem (2019) found no relations between anchoring and the Big Five.

Two possible reasons for the contradictions are (1) the use of different scales and (2) false positives (e.g., due to publication bias). Notably, different scales can explain null findings and positive findings but not negative findings and positive findings. The largest difference in scales is between the Big Five and broader HEXACO (Lee & Ashton, 2004) due to the additional factor of honesty-humility included in the HEXACO, which is represented as a facet of agreeableness in the Big Five. This leads to agreeableness correlations between the two inventories to be rather low (e.g., $r(198) = .36$; Lee et al., 2005, p. 1447). The other facets' correlations, however, overlap to a higher degree (e.g., $r(198) = .83$ for extraversion; Lee et al., 2005, p. 1447).

3.1.2. Method

In order to assess the possibility of false-positives I computed the meta-analytic effect sizes and the heterogeneity of effect sizes for all studies listed in the narrative literature review on anchoring and Big Five by Cheek and Norem (2019)[8]. Using all published effects is difficult as too little data is reported in many cases and models are heterogeneous. For example, Furnham et al. (2012) and McElroy and Dowd (2007) computed interactions between anchors (high vs.

---

[8]The considered studies are: Caputo (2014); Cheek and Norem (2019); Eroglu and Croxton (2010); Furnham et al. (2012); Jasper and Ortner (2014); McElroy and Dowd (2007); Perry and Sibley (2013); Teovanović et al. (2015).

low) and the respective scores on the Big Five but they did not report the direction of the effects that were not significant. These effects had to be excluded from the analyses. Other problematic cases are Perry and Sibley (2013), who did not use the Big Five but HEXACO model or Furnham et al. (2012) who did not report exact F-values if they were below 1 ($F = 0$ and $r = 0$ was assumed in these cases). Making various assumptions about the reported data, computation of multilevel random effects meta-analyses was possible (see https://osf.io/j24hp/ for the extracted data and https://osf.io/p48t6/ for the analysis code). Using the findings summarized by Cheek and Norem (2019) including their own findings, there are a total of 11 studies reporting 37 effects.

3.1.3. Results

Publication Bias and Questionable Research Practices. I checked for publication bias and questionable research practices via funnel plot asymmetry and p-curve analysis using all available findings. Note that funnel plots are especially useful to detect publication bias if there is a desired direction of the effect. This is not the case for Big Five moderators, which is corroborated by a nonsignificant test for asymmetry, $t(36) = -1.03$, $p = .173$, *intercept* $= -1.027$. That is, the funnel plot is symmetrical and does not indicate publication bias. However, the funnel plot of all Big Five findings (Figure 3) shows that most significant findings stem from studies with relatively large standard errors (i.e., small sample size) and p-values close to $p = .050$ indicating the possibility of false-positives.

Figure 3

Funnel plot for 37 findings on the correlation of Big Five and susceptibility to anchoring
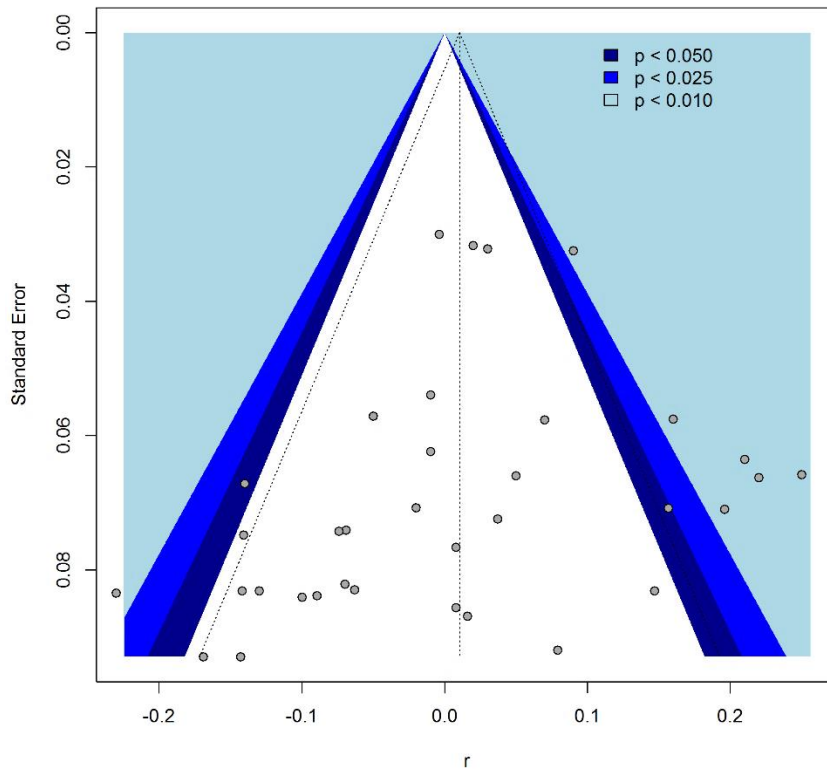
A method that is less dependent on the direction of the effects is p-curve. Among the 9 significant findings, tests for right-skewness are significant (both $p < .001$) and both tests for flatness are not significant (all $p > .700$), indicating no evidence for questionable research practices (Figure 4; see https://osf.io/p48t6/ for the code).

Figure 4

P-curve for statistically significant moderator effects of the Big Five on anchoring effects
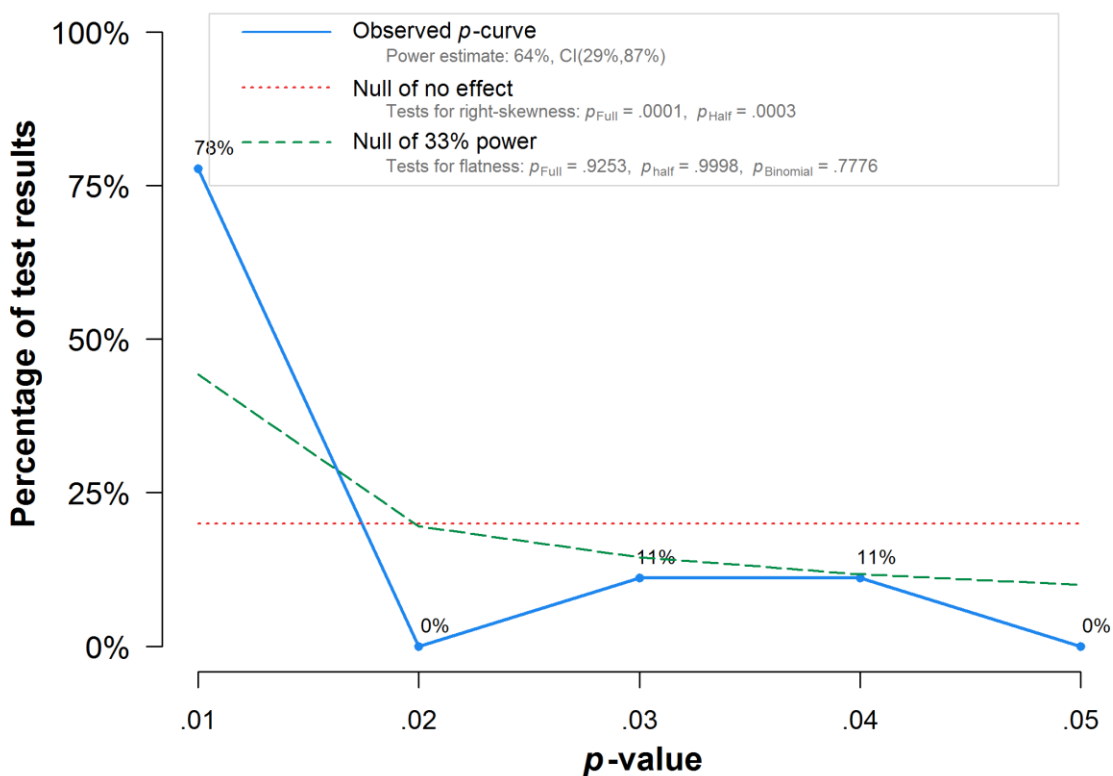
Note: The observed *p*-curve includes 9 statistically significant (*p* < .05) results, of which 7 are *p* < .025. There were 28 additional results entered but excluded from *p*-curve because they were *p* > .05.

Meta-Analytical Effect Sizes. I tested for each of the Big Five whether the meta-analytical effect size was different from zero. Notably, there was no heterogeneity for the overall effect for all studies regardless of the moderator, $Q(36) = 7.08$, $p > .999$. The total effect was $r = .013$, $p = .747$, 95% CI = [-.070, .096]. Results for the individual factors are reported in Figure 5 through Figure 9.

Figure 5

Correlations between Openness to Experience and Susceptibility to Anchoring
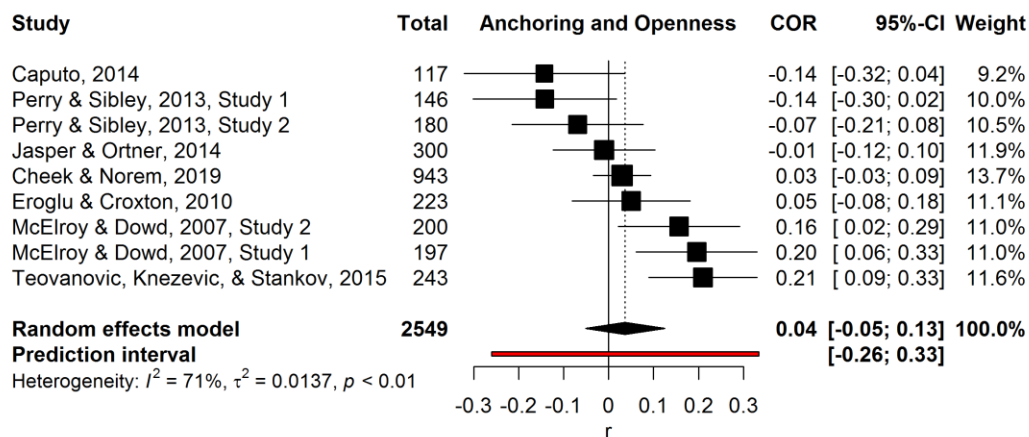
Figure 6

Correlations between Conscientiousness and Susceptibility to Anchoring
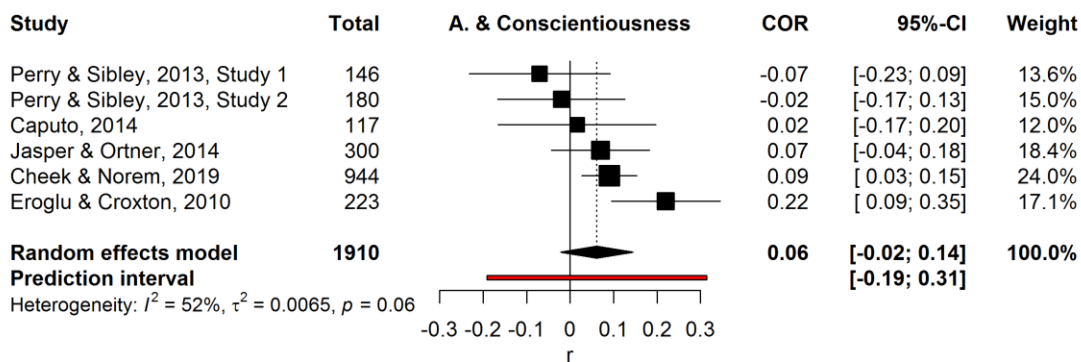
Figure 7

Correlations between Extraversion and Susceptibility to Anchoring
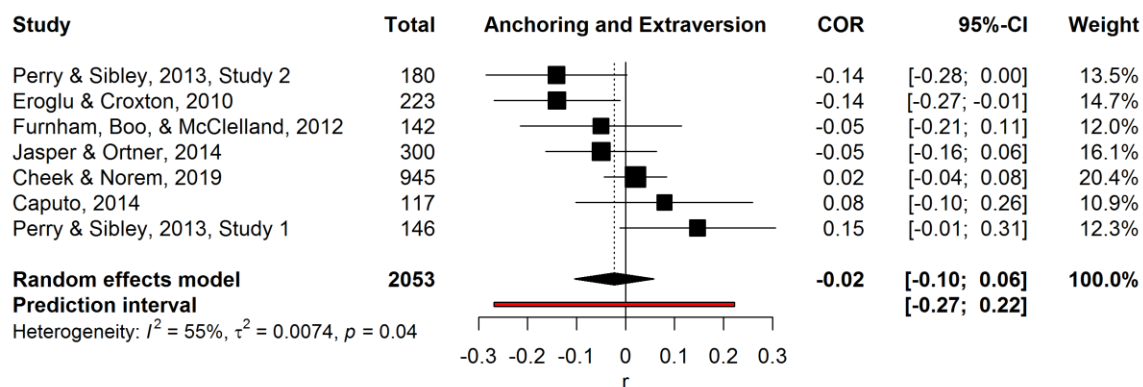
Figure 8

Correlations between Agreeableness and Susceptibility to Anchoring
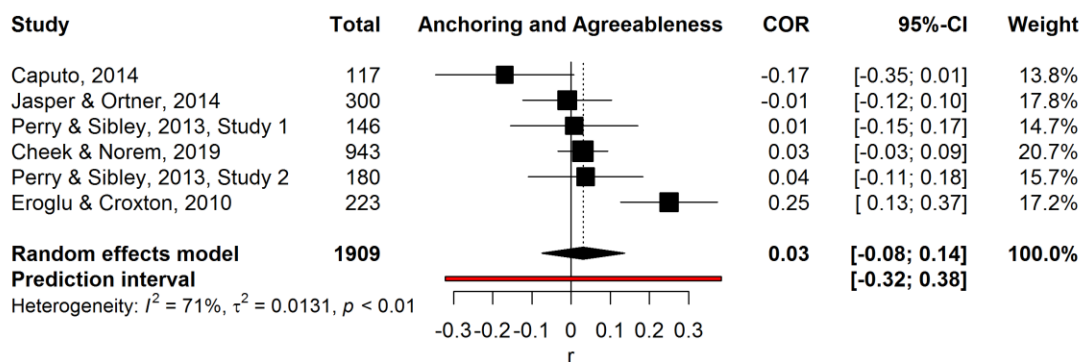
Figure 9

Correlations between Emotional Instability or Neuroticism and Susceptibility to Anchoring

Study	Total	Anchoring and Neuroticism	COR	95%-CI	Weight
Perry & Sibley, 2013, Study 1	146		-0.13	[-0.29; 0.03]	13.7%
Perry & Sibley, 2013, Study 2	180		-0.07	[-0.22; 0.07]	15.0%
Eroglu & Croxton, 2010	223		-0.01	[-0.14; 0.12]	16.4%
Caputo, 2014	117		-0.01	[-0.19; 0.17]	11.9%
Cheek & Norem, 2019	945		-0.00	[-0.07; 0.06]	24.3%
Jasper & Ortner, 2014	300		0.16	[0.05; 0.27]	18.8%
Random effects model	**1911**		**-0.00**	**[-0.08; 0.08]**	**100.0%**
Prediction interval				**[-0.25; 0.24]**	

Heterogeneity: $I^2 = 57\%$, $\tau^2 = 0.0061$, $p = 0.04$

Note. Perry and Sibley (2013) used the HEXACO dimension emotionality

3.1.4. Discussion

The results indicate very small overall correlations between the Big Five and susceptibility to anchoring. None of the pooled effects was significantly different from zero. Moreover, as this is mainly an exploratory approach where no correlation is more likely than the others, correction for alpha inflation should be applied (e.g., Bonferroni correction with $\alpha_{corrected}$ = .01) when analyzing the five personality factors with five tests. The positive or negative findings of Big Five factors correlating with the susceptibility to anchoring are thereby disqualified as likely false-positive findings. Notably, as indicated by the funnel plot and p-curve there is no reason to assume that these are due to questionable research practices or publication bias.

3.2. Meta-Analysis for Path 2: Cognitive Ability

Besides the Big Five, cognitive ability has received a lot of attention in anchoring moderator research. Different constructs related to intelligence such as SAT-scores (crystallized

intelligence) have already been correlated with anchoring strength, but the literature also includes measures of fluid intelligence (e.g., Bergman et al., 2010; Sjöberg et al., 2006). The descriptions of this path summarize the efforts by Röseler et al. (2019).

3.2.1. Theory

Röseler et al. (2019) tried to resolve the conflict between positive and null findings of cognitive ability as a moderator of anchoring by conducting a preregistered study with a paradigm based on Stanovich and West's (2008) framework for conceptualizing individual differences on heuristics and biases tasks (p. 687). According to the framework, for cognitive ability to moderate any heuristics or biases task, there must be a correlation between intelligence and a parameter that is crucial for whether an intuitive or heuristic (System 1) response or whether a deliberate or a reflected (System 2) response is given.

3.2.2. Method

Röseler et al. (2019) identified the parameter "individual's decoupling capacity to sustain override" as a correlate of intelligence (more specifically *reasoning*, Jäger & Heydasch, 2016), because Stanovich and West (2008) describe it as consisting of "probabilistic thinking skills, falsifiability tendencies, disposition to search for alternative explanations, sensitivity to contradictions, etc." (p. 687). The incremental value of considering this model is that multiple conditions must be met for cognitive ability to influence the susceptibility to anchoring. For example, participants need to detect that they have to override their heuristic response.

Although the paradigm was adapted to the model, the hypothesis could not be confirmed. That is, susceptibility to anchoring effects is *not* negatively associated with reasoning ability even if anchors were determined randomly and were thus uninformative (and participants were

pointed to the need to override the heuristic response). Instead, there was no correlation between reasoning ability and individuals' susceptibility to anchoring. Moreover, Röseler et al. (2019) found that the reliability of a series of 20 anchoring items from three different categories (quantity estimates, celebrity age estimates, willingness to pay) was very low. I will later discuss how difficulties in moderator detection could be due to low reliability of the susceptibility to anchoring. The preprint of Röseler et al. (2019) can be found in Appendix II (Article 3: Anchoring and Cognitive Ability).

3.2.3. Results and Discussion

Using the code (https://osf.io/zc34s/) provided by Röseler et al. (2019), I reproduced the finding that there is no significant overall effect across the 5 studies, $t(15) = 0.22$, $p = .827$, $g = 0.004$, 95% CI = [-0.030, 0.037], $N_{total} = [1435, 1448]$ (see also Figure 10). The first significant effect (Bergman et al., 2010) is possibly due to confounding cognitive ability with willingness to pay and the other one was revealed by an exploratory analysis by Röseler et al. (2019) and is probably a false-positive. After all, there is no theoretical reason why the susceptibility to anchoring should be moderated by cognitive ability in the quantity estimate tasks but not in the other two tasks.

Figure 10

Forest plot and prediction interval for all effects of cognitive ability on strength of anchoring effects

Study	Anchoring and Cognitive Ability	SMD	95%-CI
Bergman et al., 2010		0.48	[0.11; 0.86]
Stanovich & West, 2008, African Countries		0.16	[-1.84; 2.15]
Stanovich & West, 2008, Redwoods		0.00	[0.00; 0.00]
Furnham et al., 2012, General Intelligence A		0.09	[-0.32; 0.50]
Furnham et al., 2012, General Intelligence B		0.00	[0.00; 0.01]
Furnham et al., 2012, General Intelligence C		0.00	[0.00; 0.01]
Furnham et al., 2012, General Intelligence D		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning A		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning B		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning C		0.09	[-0.32; 0.50]
Furnham et al., 2012, Verbal Reasoning D		0.00	[0.00; 0.01]
Teovanovic, 2019, Raven's Matrices		0.06	[-0.03; 0.15]
Teovanovic, 2015, Raven's Matrices		0.06	[-0.18; 0.30]
Röseler et al., 2019, Quantity Estimates		0.49	[0.24; 0.75]
Röseler et al., 2019, Age Estimates		0.10	[-0.16; 0.36]
Röseler et al., 2019, Price Estimates		0.02	[-0.22; 0.26]

-1   -0.5   0   0.5   1

Hedge's g

3.3. Meta-Analyses for Path 3: Cognitive Reflection and Impulsiveness

3.3.1. Theory

Using the TOTE-model that is based on insufficient adjustment, cognitive ability can be hypothesized to reflect a longer adjustment process. For example, the construct *cognitive reflection* (Frederick, 2005) with its items such as "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" (p. 27) requires respondents to reject the answer that quickly springs to mind. So does giving an estimate that is far away from the anchor if the insufficient adjustment account is valid: The "test" component of TOTE (test, operate, test, exit) has to be negated more often in order to make many adjustments and arrive at a value that is unbiased from the anchor.

In a similar way, impulsiveness should be associated with a shorter adjustment process. However, impulsiveness (or self-control, which is moderately to highly correlated with, self-

control or impulse control should lead to a rejection of the first intuitive answer and thus impulsiveness, e.g., r[521] = .64, Nebioglu et al., 2012) is not a strong predictor of cognitive reflection, r(338) = -.19, and r(192) = -.21 (Littrell et al., 2020, pp. 5 and 7, respectively).

3.3.2. Method

We conducted meta-analyses for all available reports of moderation analyses (i.e., from published reports, preprints, and preregistrations) for cognitive reflection and impulsiveness (search terms: anchoring AND cognitive reflection, webofknowledge.com, all databases).

3.3.3. Results

Cognitive Reflection Task (CRT). Note that CRT and cognitive ability should not lead to very different results as CRT scores consist to a large part of cognitive ability. For example, their correlation in the study by Bergman et al. (2010) is r(114) = .620, p < .001. With the reliability of CRT being α < .7 in many cases (e.g., α = .40, Teovanović, 2019, p. 16; α = .66, Röseler, Schütz, Dolling et al., 2020), there is not much variance left to be explained by constructs other than intelligence. Moderations of anchoring effects by CRT scores were reported in several articles (Bergman et al., 2010; Czerwonka, 2017; Meub et al., 2013; Oechssler et al., 2009; Röseler, Schütz, Dolling et al., 2020; Teovanović et al., 2015; Teovanović, 2019; Welsh et al., 2011, 2014). Similar to other moderators, some reports did not include enough data to include them in the meta-analysis or correlations had to be computed indirectly and by making several assumptions. For example, Oechssler et al. (2009) reported only the sample size and the fact that the p-value was larger than .05, which is why the study has not been included in the meta-analysis. Another problem is that in some studies the CRT score was split into low (0 or 1 correct

answers) and high (2 or 3 correct answers) groups, which reduces power and the effect size (e.g., McClelland et al., 2015).

There were 10 effects in total (data and analysis script are available online, https://osf.io/5utwj/) of which one was significantly smaller than 0. A random effects model revealed no effect of cognitive reflection on the susceptibility to anchoring, $z(9) = -0.28$, $p = .779$, $r = -.010$, 95% CI = [-.082, .061], $N_{total} = 1242$. Correlations were homogeneous, $Q(9) = 9.80$, $p = .367$. To sum up, there is overwhelming evidence for cognitive reflection to be independent of the strength of anchoring, which is consistent with the findings on moderator effects of intelligence on anchoring described above.

Figure 11

Forest plot of effects of cognitive reflection on the susceptibility to anchoring
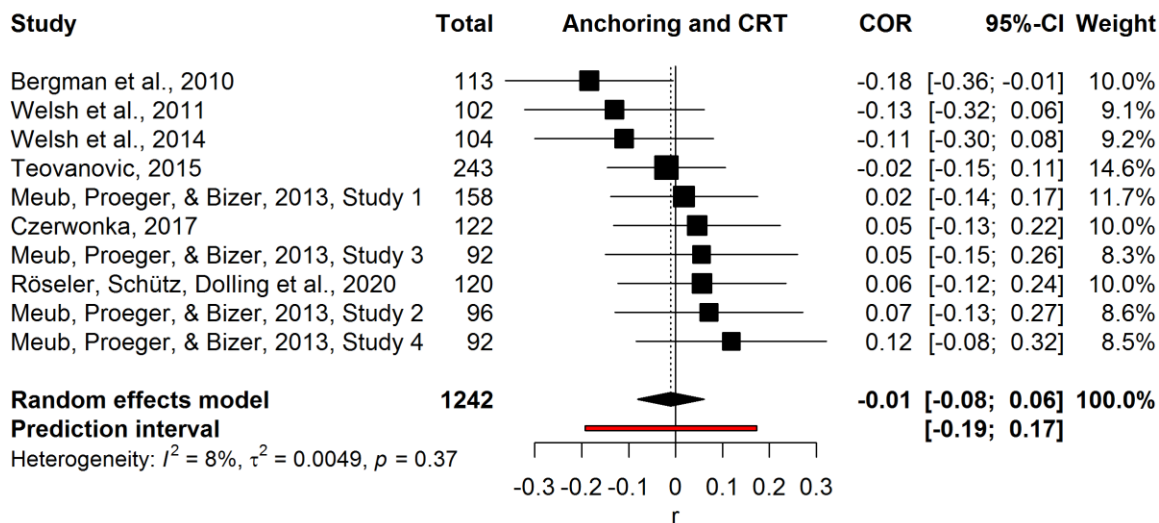
Impulsiveness. Röseler et al. (2019) reported a null-correlation between anchoring and the Brief Self-Control Scale: "$r(223) = -.02$, $p = .753$" (p. 15). In a registered report, Röseler, Schütz, Baumeister, and Starker (2020) reported that "participants' mean anchor susceptibility

was not associated with low self-control, $r(648) = -.013$, $p = .739$" (p. 6). Using the data

provided by Röseler, Schütz, Dolling et al. (2020), the correlation between the mean adjustment

score and impulsiveness measured with the BIS-11 (e.g., Stanford et al., 2009) can be computed,

$r(120) = .155$, $p = .088$ (and reversed, because impulsiveness is the opposite of self-control). No

other reports of effects of trait self-control on anchoring could be found. The combined

correlation is not significant, $z = -0.90$, $p = .370$, $r = -.039$, 95% CI $= [-.124, .046]$, $N_{total} = 997$.

Effects were homogenous, $Q(2) = 1.79$, $p = .409$.

Figure 12

Forest plot for effects of self-control on strength of anchoring effects
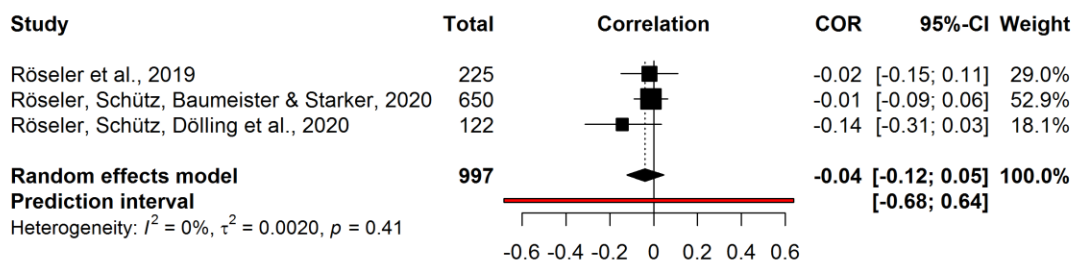
3.3.4. Discussion

There is no evidence for moderator effects of cognitive reflection or impulsiveness and

self-control. Mean effect size estimates are close to $r = 0$ with very large sample sizes ($N = 1242$

and 997, respectively) and low heterogeneity.

3.4. Explaining Contradictory Findings in Anchoring Moderator Research

Little has changed since Furnham and Boo (2011) claimed that "researchers have failed

to identify any … trait variables that have a systematic and explicable effect on anchored

decisions" (p. 40) and no solution seems in sight. Table 3 summarizes the results on personality

trait moderators. Different arguments have been made why this could be the case. The most

promising argument is the hitherto unnoticed reliability problem of the susceptibility to anchoring (Röseler et al., 2019).

Table 3

Overview of meta-analytical effects of several personality moderators on anchoring effects strength

Moderator	r	95% CI	N	Reasons for contradictions
Neuroticism	.00	[-.08, .08]	1911	False positives
Extraversion	-.03	[-.12, .05]	2053	False positives
Openness	.04	[-.05, .13]	2549	False positives
Agreeableness	.03	[-.08, .14]	1909	False positives
Conscientiousness	.06	[-.02, .14]	1910	False positives
Cognitive Ability	.00	[-.01, .02]	1435	False positives
Cognitive Reflection	-.01	[-.08, .06]	1242	No contradictions
Self-Control	-.04	[-.12, .05]	997	No contradictions

Note. CI = Confidence Interval.

3.4.1. Computation of Susceptibility to Anchoring

There are many ways to compute an individual's susceptibility to anchoring. For example, one can simply use the estimate for a given question and anchor, or compute the relative absolute adjustment, that is, the absolute difference between the anchor and the estimate divided by the true value (if a true value for the question exists) or an unanchored estimate (if an unanchored for the question exists). In some cases, authors used the difference between anchor and estimate. Cheek and Norem (2018, p. 2) argue that the use of differences between anchors and estimates is an inappropriate measure of susceptibility to anchoring. They recommend using *absolute* differences, instead. Röseler et al. (2019, p. 14) followed the recommendation and did not find it to alter the correlation between cognitive ability and the susceptibility to anchoring.

In paradigms with within-subjects manipulation of the anchors (i.e., the correct value is the same for two items but the anchors are above and below the value; e.g., Röseler, Schütz, Dolling et al., 2020) one can also compute the difference between the two respective estimates to

measure the susceptibility to anchoring. However, the use of difference scores is often (and necessarily) associated with low reliability regardless of the construct under investigation (Peter et al., 1993). Röseler, Schütz, Dolling et al. (2020, Table 2) compared the reliabilities of absolute adjustment scores with difference scores and found no differences in the inter-item correlations, $r(121) = .052$, and $r(121) = -.086$, respectively.

Another area of research that is underrated for the computation of anchoring scores is research on the association between anchor extremeness and anchoring effect strength. The knowledge that doubling the anchor's distance from the true value does not double the effect size (Chapman & Johnson, 1994; Thorsteinson, 2011; Wegener et al., 2001) has not yet been integrated in the computation of anchoring scores. Moreover, there is no consensus with regards to the focal elements of the anchoring paradigm such as (a) reference point and (b) distances of anchors. (a) For example, if people have to estimate the number of words in this section, the reference value can be the true number but also the unanchored estimate of the same or another sample. (b) In studies where multiple items have the same true value and anchors are manipulated, it is unclear whether they should be symmetrical or asymmetrical and whether they should be absolute values such as ±50 units (e.g., Röseler et al., 2019) on the respective scale or relative values such as ±50% (e.g., Teovanović, 2019). Ideally, researchers should know the relationship between the size of the anchor and anchor effect strength, include a parameter for the individual and compute the individual susceptibility to anchoring based on that relationship. However, such an approach does not yet exist.

3.4.2. Publication Bias and the Use of Questionable Research Practices

Generally, publication bias and the use of questionable research practices such as p-hacking (Simmons et al., 2011) leads to a skewed picture of effect sizes in the literature.

Although cases such as incidental anchoring or effects of ego depletion on anchoring are very likely cases of p-hacked results as discussed above, funnel plots and p-curve analyses of moderator findings on the Big Five, cognitive ability, cognitive reflection, and impulsiveness are not consistent with this accusation. Instead, the majority of findings are null-findings. The outliers are much more likely to be false-positives than the products of questionable research practices.

3.4.3. Low Reliability of Personality Trait Scales

Second, low reliability of short versions of personality trait scales used in some of the research might lead to null results (Teovanovic, 2019, p. 17). Overwhelming evidence of research using reliable scales and at the same time not finding significant correlations with traits (Cheek & Norem, 2019; Furnham & Boo, 2011; Röseler et al., 2019; Röseler, Schütz, Baumeister, & Starker, 2020) stands against this argument. Moreover, the problem of low reliability can be neglected with very large sample sizes. That is, even for scales with relatively low reliability such as CRT (Frederick, 2005) sample sizes of $N > 1000$ could even reveal very small correlations.

3.4.4. Low Reliability of Susceptibility to Anchoring

Lastly, reliability of the measured susceptibility to anchoring might have been too low (e.g., $\alpha < .1$). This argument is the most serious one because measurements with this low reliabilities not only attenuate correlations but make them impossible to occur. Although every psychologist is taught the necessity of reliability, assessing it for instruments is not a common practice: Some tasks have been used for many years although the reliability is low (e.g., dot-probe task, Parsons et al., 2018; FMRi, Elliott et al., 2020; attentional-control, Bastian et al.,

2020) or had been unknown (e.g., cognitive reflection task, Primi et al., 2016). Anchoring is not an exception to this problem: In order to assess this concern, Röseler et al. (2019, Table 3) analyzed reported reliabilities of several anchoring tasks and computed reliability for tasks with public datasets. The reliabilities are generally low and average correlations are close to zero (e.g., $r = .03$; Cheek & Norem, 2019).

3.5. Solving the Reliability Issue of Susceptibility to Anchoring

Due to the low reliability of anchoring scores, personality moderator effects cannot be found even if they existed. A serious concern is the non-existence of the trait susceptibility to anchoring (as suggested by Röseler, Schütz, Baumeister, and Starker, 2020). Note that the failure of reliable assessment does not imply a construct's non-existence. Researchers have not even tried to reliably assess susceptibility to anchoring. Note also that the successful reliable assessment of anchoring scores is merely a necessary condition of the construct's existence and construct validation is still needed. Thus, different methods of assessing the susceptibility to anchoring need to be tested next. Reliable estimates need to be validated through establishing construct validity. If none of them succeeds, researchers have reasons to consider abandoning personality moderator research in the realm of anchoring effects.

A TOTE-model inspired approach that focuses on the *number of adjustment steps* (i.e., how often people adjust their estimate before submitting it) is provided by Frech et al. (2020). Röseler, Schütz, Dolling et al. (2020) analyzed the data from Frech et al. (2020) and found that the number of adjustment steps can indeed be a reliable parameters of anchoring ($\alpha = .69$, three items, $N = 146$; see https://osf.io/yd7c8/ for the analysis script). They adapted the original paradigm (which was used for the context of price negotiations) to fit classical anchoring items that do not involve social aspects, used high and low anchors, a digital version, and more items

to test if the finding about the reliable number of adjustment steps can be generalized. The number of adjustment steps was successfully measured with relatively high reliability ($\alpha = .749$, 12 items, $N = 123$).

Moreover, a paradigm in which the number of adjustment steps and the width of the adjustment steps can be measured separately allows for more specific hypotheses. As specified above, cognitive reflection and impulsiveness were hypothesized to be associated with the number of adjustment steps (and thereby the total adjustment). Using the total adjustment as the dependent variable when the adjustment step size contains much error variance will thus render the total adjustment less reliable and thus less adequate for the hypothesis test.

In the study by Röseler, Schütz, Dolling et al. (2020), the number of adjustment steps could be measured reliably, however, the correlations with cognitive reflection and impulsiveness were not as hypothesized but opposite (cognitive reflection; $r[121] = -.218$) or zero (impulsiveness; $r[121] = .111$). It is as yet unclear what personality traits are associated with the number of adjustment steps and—in the worst case for personality science—the number merely reflects the tendency to make many (or few) clicks with a mouse.

To sum up, after more than one decade in anchoring moderator research, identifying reliable parameters of the susceptibility to anchoring is currently the only way to conduct methodologically sound (personality) moderator research. Having a reliable estimate of some sort of susceptibility to anchoring is the necessary condition for personality moderator research. More parameters need to be identified and more paradigms need to be developed to measure these parameters. The preprint of Röseler, Schütz, Dolling et al. (2020) that develops the stepwise-adjustment paradigm (STAP) to reliably assess susceptibility to anchoring can be found in Appendix II (Article 4: Reliable Measurement of Susceptibility to Anchoring).

4.   Conclusion

Contradictions and replication failures have rarely been addressed in anchoring research. They complicate inductive theory building, that is, developing explanations of anchoring based on published findings. In order to facilitate theory building, I described and provided explanations for the contradictions in (1) findings about subliminal anchoring, (2) the role of ego depletion and anchoring, and most thoroughly (3) findings about personality moderators of anchoring. I have shown that the contradictions in cases (1) and (2) emerged due to questionable research practices and that (3) the plethora of null-findings in personality moderator research can be explained by the lack of a paradigm to reliably assess individual differences in the susceptibility to anchoring. Anchoring models such as the TOTE model of insufficient anchoring have proven to provide the strongest basis for solving the reliability problem (Röseler, Schütz, Dolling et al., 2020). The successful creation of a paradigm with which the number of adjustment steps can be reliably assessed provides an important basis for methodologically sound personality trait moderator research in anchoring. The validity of the reliable scores has yet to be investigated.

4.1. False-Positives and Questionable Research Practices in Anchoring Research

Currently, the most promising strategy to resolve contradictions between published studies seems to be replicating the contradictory studies. In many cases, the results and thereby the contradiction cannot be replicated, making false-positives a simple explanation: As I have shown, this procedure worked for Ego Depletion (Röseler, Schütz, Baumeister, & Starker, 2020), Big Five (Cheek & Norem, 2019), Cognitive Ability (Röseler et al., 2019), and subliminal anchoring (Röseler, Schütz, Blank et al., 2020). Although the anchoring effect is certainly large and robust (see also Appendix I: The Size of Anchoring Effects), anchoring research in general is

definitely not less prone to questionable research practices than other fields. I have presented p-curves for some subfields, which show the absence of evidential value and explain failed subsequent replication attempts (e.g., the effect of ego depletion on anchoring or subliminal anchoring; see also Shanks et al., 2020). However, in the field of personality moderator research, few significant results have been found. Therefore, it is much more likely that false-positives have created the seemingly contradictory picture. I have conducted meta-analyses for eight personality moderators that revealed overall null-effects and absence of questionable research practices.

4.2. Limitations and Future Research

In psychology, a claim of the nonexistence of something is much harder to test than a claim of the existence of something (e.g., Ferguson & Heene, 2012). In the case of personality moderators of anchoring, the evidence is clearly in favor of null-correlations for eight of the most prominent personality moderators of anchoring (Big Five, cognitive ability, cognitive reflection, and self-control / impulsivity). Future research needs to expand upon this in two ways. (a) My literature research that yielded the findings were unsystematic because I did not always provide search terms to exactly reproduce the sample of research included in the meta-analyses. For example, searching for "anchoring AND cognitive ability" would not reveal some of the included papers (e.g., Teovanović et al., 2015). I encourage other researchers to expand the analyses and to reproduce the results that are available online (https://osf.io/9gju4/). (b) There might be studies on other moderators or unpublished studies. Although I found no evidence for publication bias, unpublished findings would contribute to the precision of the effect size estimates. Literature searches among preprints (https://osf.io/preprints/) and ongoing studies or unpublished data (https://osf.io/registries) have not yielded any results. However, few of the

studies that were included in my meta-analyses had been preregistered (cf. Cheek & Norem, 2019).

The search for reliable and theoretically sound anchoring scores has just begun and already yielded one parameter (i.e., the number of adjustment steps). Future research needs to validate the existing parameter or find new ones. If this proves difficult, the possibility of susceptibility to anchoring effects not being a personality trait may need to be considered. A lack of individual differences would be consistent with scale distortion accounts (e.g., Mochon & Frederick, 2013), which do not overlap with individual differences at all.

Although the anchoring effect itself is robust and has been replicated many times (e.g., Klein et al., 2018; see also Appendix I: The Size of Anchoring Effects), more moderator effects and key findings for theories need to be replicated to guarantee a solid base of evidence on which existing and future theories are built. I suggest adversarial collaborations (i.e., advocates of different theories working together and conducting critical tests between models; Cowan et al., 2020) in order to integrate existing models, find more contradictory hypotheses between the models, or create models with more empirical content. After all, there are still many untested hypotheses derivable from the proposed models and testing them is crucial to advance the research field.

Finally, scientific progress will also be difficult if researchers continue to ignore basic methodological facts: I have summarized more than one decade of research on personality moderators of anchoring that has corroborated with overwhelming evidence what should have been clear from the beginning: *Correlating something with an unreliable score will necessarily yield a null correlation*.

Appendix I: The Size of Anchoring Effects

Knowledge about effect sizes is crucial for informative research because sample size planning hinges on this hard to assess parameter. Recommendations about required sample sizes to find anchoring effects for items such as Mahatma Gandhi's age when he died (anchors: 64, 79) are 20 participants per group (e.g., Bahník et al., 2017, p. 224), which correspond to a minimum $d = [1.156, 1.531]$ for power of $1-\beta = [.80, .95]$. In order to test whether this recommendation holds for a variety of anchoring effects and in order to facilitate future anchoring research in terms of effect size estimation and power analyses, I provide a brief meta-analysis of the reported studies' anchoring effects. Combining the studies allows for a large sample size, assessment of the heterogeneity between anchoring effects and effect sizes estimates that are more precise. Furthermore, it allows for a strict test of the claims I make in the Open Science Statement.

There were 24 effects in the studies that are included in this work and that feature anchoring effects, (Röseler et al., 2019; Röseler, Schütz, Baumeister, & Starker, 2020; Röseler, Schütz, Dolling et al., 2020). 10 of them were administered within-subjects, only and are a mix of random and informative anchors, which have been found to have effects of different sizes (Röseler et al., 2019, pp. 13-14) and 14 of them were experimenter provided. I computed the relative distance between the high and low anchors and the true value for each item (i.e., anchor minus true value divided by anchor), which was between .13 (age estimate for 53 years old celebrities) and .88 (weight of a cat). Mean relative anchor distances also differed between studies and was thus confounded sample size (and standard error). Note that the relative distance between the anchor and the true value is an important moderator of anchoring effect strength (e.g., Chapman & Johnson, 1994).

Test for Questionable Research Practices

The reported effects were subjected to the same critical standard as the moderators, that is, I report the funnel plot and p-curve analysis results (analysis code and data available at https://osf.io/n9cy6/). Funnel plots are commonly used to determine publication bias. As all effects are reported, the funnel plot for the anchoring effects should be symmetrical. There is no symmetry of effect sizes as visible in Figure 13 and indicated by an Egger's test, $t(22) = 6.08$, $p < .001$.

Figure 13

Funnel plot of the 24 anchoring effects from the three studies included in this research
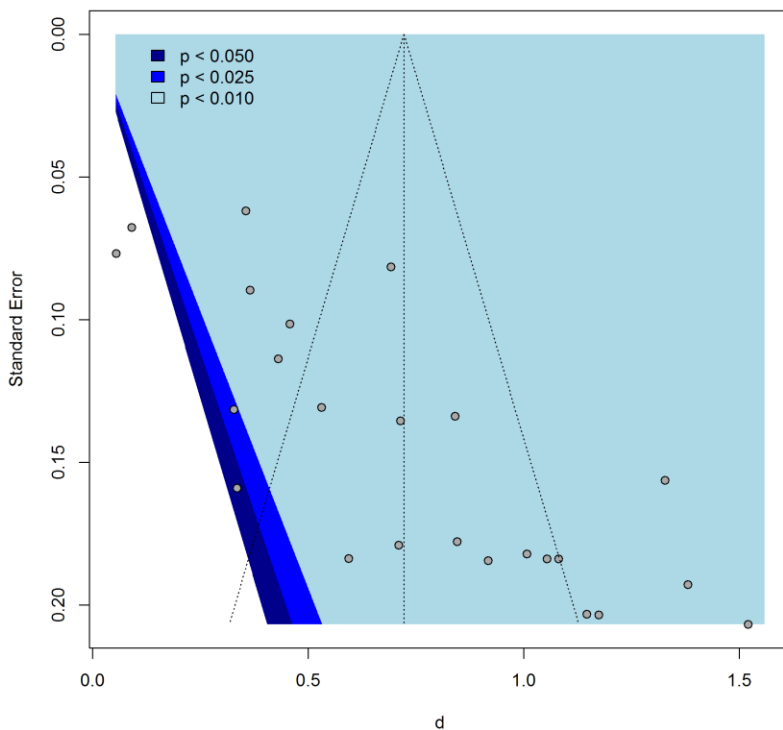
In order to explain the asymmetry of the funnel plot, exploratory analyses were conducted. They revealed that the standard error (sample size) and the relative anchor distance are confounded between studies. That is, the sample size was smaller for the studies which

featured smaller relative anchor distances (see Figure 14). Moreover, p-curve analysis revealed presence of evidential value (both tests for right-skewness $p < .001$), no inadequacy or absence of evidential value (all tests for flatness $p > .999$), estimated power of $1-\beta = .99$, 95% CI = [.99, .99], and an estimated effect size of $d = 0.726$.

Figure 14

Relative anchor distances and anchoring effects in the reported studies
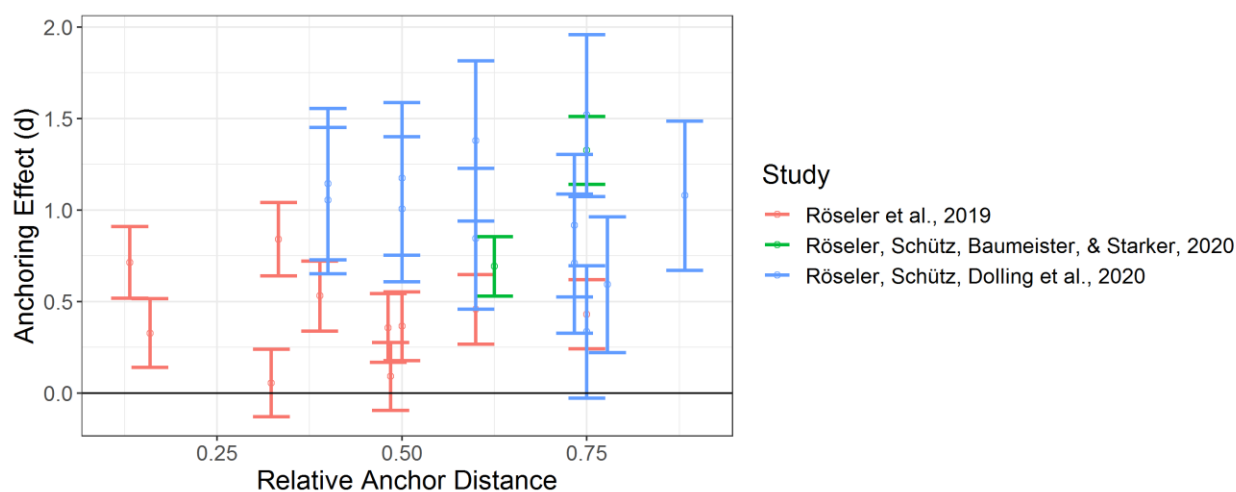
Meta-Analytical Effect Size and Heterogeneity

I conducted a multilevel random-effects model meta-analysis in order to estimate the effect size with consideration of the hierarchical structure (24 effects nested in 3 studies). The mean effect size was very similar to the estimate from the p-curve and significantly different from 0, $t(23) = 3.56$, $p = .002$, $d = 0.725$, 95% CI = [0.304, 1.147], $N_{total} = 997$. Heterogeneity was low, $Q(23) = 29.22$, $p = .173$.

Figure 15

Plot of Anchoring Effects Featured in the Reported Studies

Source	SMD (95% CI)
Quantity 50	0.46 [0.26; 0.66]
Quantity 40	0.43 [0.21; 0.65]
Quantity 60	0.37 [0.19; 0.54]
Age 69	0.33 [0.07; 0.59]
Age 53	0.71 [0.45; 0.98]
Age 36	0.53 [0.27; 0.79]
Age 30	0.84 [0.58; 1.10]
WTP pen	0.05 [-0.10; 0.21]
WTP apple	0.36 [0.23; 0.48]
WTP keyboard	0.09 [-0.04; 0.22]
Distance Warsow to Liverpool	0.69 [0.53; 0.85]
Life expectancy of a porcupine	1.33 [1.02; 1.63]
Weight of a rabbit	0.59 [0.23; 0.95]
Weight of a cat	1.08 [0.72; 1.44]
Weight of a fox	1.52 [1.11; 1.93]
Weight of a racoon	0.34 [0.02; 0.65]
Weight of a badger	0.92 [0.56; 1.28]
Weight of a pelican	0.71 [0.36; 1.06]
Weight of a beaver	0.85 [0.50; 1.19]
Weight of a deer	1.38 [1.00; 1.76]
Weight of a wolf	1.01 [0.65; 1.36]
Weight of a goat	1.17 [0.78; 1.57]
Weight of a hyena	1.05 [0.69; 1.41]
Weight of a warthog	1.15 [0.75; 1.54]

Recommendations for a priori Power Analyses in Anchoring Research

As discussed in the introduction, anchoring effects exceed commonly found effect sizes by far, making research on them much less costly because relatively few participants are needed. However, when it comes to precise estimates that are needed for a priori power analyses (and the determination of the necessary sample size), even a collection of 24 effects and a sample size of almost 1000 individuals do not yield a definite answer. First, anchoring effect sizes depend on multiple design features, such as the extremeness of anchors (e.g., Chapman & Johnson, 1994) and the expertise of the respondents (e.g., Frech et al., 2019). Second, even minor design choices have been shown to have a non-negligible impact on effect sizes (Landy et al., 2020) especially if the effects are large (Olsson-Collentine et al., 2020) as is the case for anchoring. For precise power analyses, I recommend choosing a mean effect size of anchoring items that closely resemble the used items and considering factors such as prior information or how informative the anchors are (Turner & Schley, 2016). Beware that heterogeneity represents a lack of understanding of moderators and design choices (e.g., Linden & Hönekopp, in press). Although effect sizes are difficult to find in some cases, availability of data in anchoring research has increased over the last years. Even using the lower confidence interval bound for effect size estimation in power analyses yields achievable sample sizes in anchoring research: For example, Röseler, Schütz, Dolling et al.'s (2020) lower 95% CI effect size bound for animal weight estimates is $d = 0.76$, which would require a total sample size of $N = 78$ for 95% power.

Appendix II: Research Articles

This appendix includes the four research articles referenced in the text and one additional research article on the social perception of trait self-control. Whether the published or the submitted version of the respective article is included, depends on the open access policies from respective publisher.

Article 1: Anchoring and Ego Depletion

The following manuscript is the submitted version of a manuscript that has been published at the Journal of Experimental Social Psychology. Note that due to the manuscript being a *registered report*, there are no results and discussions in the submitted version.

Röseler, L., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce judgment adjustment for both internally and externally generated anchors? *Journal of Experimental Social Psychology*, *87*, 103942. https://doi.org/10.1016/j.jesp.2019.103942

Does Ego Depletion Reduce Judgment Adjustment for Both Internally and Externally Generated Anchors?

Lukas Röseler

Harz University of Applied Sciences and University of Bamberg

Astrid Schütz

University of Bamberg

Roy F. Baumeister

University of Queensland

Ulrike Starker

Harz University of Applied Sciences

Abstract

Ego depletion is a state in which people prefer to avoid mental effort, therefore possibly leading to increased reliance on heuristics. Effortful thinking has been shown to help reduce anchoring effects, in which people form social judgments by adjusting from an initial value (the anchor). We therefore predicted that ego depletion would reduce the amount of adjustment from an initial anchor, leaving the final judgment relatively close to the anchor value. In contrast to previous research by Banker et al. (2017), we exclude alternative explanations, such as social mechanisms. In particular, we test adjustment from internally generated anchors. We theorize that judgment adjustment processes are the same for both internal and external cues.

Keywords: self-regulation; self-control; depletion; anchoring, pre-registration

The effect of ego depletion on susceptibility to external and internal cues

Ego depletion refers to a state in which a person's effortful control has been strained and the person prefers effortless to effortful thinking. This state typically occurs after a person has engaged in effortful cognitive activity, for example by suppressing one's emotions or making many decisions (Vohs & Heatherton, 2000). Resisting *anchoring effects* (e.g., not letting a numerical value bias one's estimation) can be considered effortful thinking. Based on these premises we hypothesized that ego depletion should lead to stronger anchoring effects.

Anchoring and adjustment is a heuristic process proposed by Tversky and Kahneman (1974). When trying to estimate an unknown value, the person may start from a relevant known value (the anchor) and then make adjustments. Much research has found that there is a general tendency for the adjustment to be inadequate, so that the judgment remains closer to the anchor than it should be. This process emerged as relevant to ego depletion in an investigation by Banker et al. (2017). Preliminary studies using the Dictator Game, in which one person divides a cash reward between self and another, found that depleted Dictators kept more money for themselves, which could indicate selfishness (selfishness hypothesis). However, when the procedure was refined so that the money was first given to the other person and the dictator could then take as much as he or she wanted, depleted dictators were exceptionally unselfish. The implication was that depleted dictators tended to leave the money wherever it was. If it had initially been given to them, they kept more of it, but if it had initially been given to the other player, they left it there. Banker et al. (2017) concluded that ego depletion leads to stronger influence of *external cues* (sticky anchor hypothesis of ego depletion). This is consistent with other evidence that ego depletion creates a general attitude of passivity, in which the self

becomes less likely to take active control and engage in proactive mental activity (Vonasch, Vohs, Ghosh, & Baumeister, 2017).

We sought to extend that research in two key respects. First, anchoring effects in that study by Banker et al. (2017) were not operationalized in the usual manner, which involves having the person generate the anchor internally and arbitrarily (e.g., by using the numbers from one's birthday) and estimating whether some parameter is higher or lower. Instead, an amount of cash was allocated to self or other. The social and monetary aspect may have introduced motivational factors that interacted with the ego depleted state, such as attitudes toward the partner, guilt or pride over one's allocation, and wishing to make a good impression on the experimenter.

Thus, although Banker et al. (2017) interpreted their findings in terms of anchoring, two large questions remained because they did not use a standard anchoring procedure. The first is whether the interpersonal context was responsible for the effects: Participants were not simply estimating a parameter, as in much anchoring research, but instead were re-allocating money between themselves and another person. Obviously, the interpersonal aspect introduces concerns beyond simple judgment and estimation, such as fairness, moral obligation, and cooperation. The second is whether Banker et al.'s findings were specific to externally generated anchors, which again may introduce extraneous factors (e.g., submission to external dictates), as opposed to conforming to the classic anchoring work, which relies on internally generated anchors. Hence Banker's findings raise (rather than obviate) the importance of testing whether ego depletion would increase the anchoring effect in the standard paradigm involving one person estimating a parameter in relation to an internally generated anchor.

In the dictator game paradigm studied by Banker et al. (2017), the initial allocations were determined by the experimenter (i.e., assigning all the money initially to either the participant/dictator or to the non-dictator partner). The sticky anchor hypothesis of ego depletion as described by Banker et al. (2017) builds upon the theory that "explicit self-reports reflect automatic associations to a greater extent when people do not have the motivation or the cognitive capacity to retrieve additional information from memory" (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005, p. 1380). However, so far the hypothesis has been tested only for external cues (Banker et al., 2017, Haan & van Veldhuizen, 2015). We will extend the research by using *internal* cues. In sum, we will choose a classical anchoring paradigm to test whether depletion has an effect on anchoring while excluding social mechanisms and including internal cues in the form of self-generated anchors.

The intended contribution of this investigation is as follows: We build upon Banker et al. (2017), who considered only one mechanism of anchoring, namely the selective accessibility model (p. 2), and we will briefly show that the observed effects would be predicted on the basis of other mechanisms, such as anchoring and adjustment (e.g., Epley & Gilovich, 2001), too. Second, we examine ego depletion's effects on susceptibility to anchoring effects on the basis of a classic anchoring paradigm. This way we will provide a cleaner test of the prediction that ego depletion affects anchoring. Third, we include self-generated anchors, which we consider to be *internal cues*. We believe it to be important to corroborate the sticky anchor hypothesis for internal cues: if self-generated anchors, i.e. internal cues are affected by ego depletion, too, this would open new research avenues and hypotheses, for example it would predict stronger IAT-patterns for depleted participants. On the other hand, if the effect turns out to be limited to the interpersonal context studied by Banker et al., our broad hypothesis would be falsified. That

would suggest that the role of ego depletion would be to alter susceptibility to socially motivational factors rather than simply decreasing the internal process of effortful adjustment.

Finally, by combining two lines of research we aim to learn more about how both of them (i.e. ego depletion and anchoring work. Note that both fields are penetrated by theoretical questions: Anchoring effects – although there existence is undisputed – are subject to numerous theories which seem to be exchanged arbitrarily in order to explain new findings. Ego depletion is subject to theoretical and methodological criticism. There is uncertainty about its existence despite a plethora of studies (e.g., Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2018; Baumeister, Tice, & Vohs, 2018). If we can confirm our hypothesis we will provide evidence for the existence of the ego depletion effect and add enlarge the body of pre-registered ego depletion research (e.g., Dang, Liu, Liu, & Mao, 2017). If the hypothesis is not confirmed, the present design will, however, not permit us to conclude that the ego depletion effect does not exist, but only that if ego depletion does exist, it does not alter the anchoring and adjustment process.

Derivation of hypotheses

There are several competing theories that seek to account for anchoring effects, and more than one may be correct (e.g., Simmons, LeBoeuf, & Nelson, 2010). Apart from selective accessibility as a possible mechanism predicting anchoring effects (p. 2), there is a host of other mechanisms that have been proposed (e.g., Tversky & Kahneman, 1974) and each mechanism has been supported by some findings while not being able to account for others (e.g., the proposed mechanism *numeric priming* does not account for the finding that anchoring the number of calories of French fries does not anchor how many grams of fat they contain, Frederick & Mochon, 2012). As the present study will focus on the role of ego depletion in

anchoring, we will briefly discuss how ego depletion may affect each of the proposed

mechanisms.

Theories and mechanisms that are put forward to explain anchoring effects include (1)

confirmatory testing, (2) semantic priming, (3) numeric priming, (4) insufficient adjustment, and

(5) scale distortion. The first two of these, (1) confirmatory testing and (2) semantic priming, are

rooted in the selective accessibility model (Mussweiler & Strack, 1999) and are what Banker et

al. (2017) featured in their research: Concerning confirmatory testing, "depletion may impair the

individual's ability to consider options counter to the suggestion" (Wheeler et al., 2007, as cited

in Banker et al., 2017) and may thereby lead to stronger anchoring effects. In this argument,

depletion leads people to engage in more confirmatory information processing and less reasoning

(Fischer, Greitemeyer, & Frey, 2008; Schmeichel, Vohs, & Baumeister, 2003). The argument

dovetails with findings from research on anchoring: *Considering the opposite*, that is taking into

account arguments that are inconsistent with the anchor, weakens anchoring effects (Mussweiler,

Strack, & Pfeiffer, 2000).

Besides confirmatory testing, semantic (2) and numeric priming (3) have been proposed

as mechanisms responsible for anchoring effects. Depleted participants may be more susceptible

to all priming effects (e.g., Bertrams, Baumeister, Englert, & Furley, 2015). Also "priming

selective accessibility (a) impacts the anchoring bias independently of the type of anchor and (b)

interacts with effort in the same way across both sources of anchors " (Chaxel, 2014, abstract).

Originally, anchoring effects were referred to as (4) insufficient adjustment (Tversky

& Kahneman, 1974). Epley and Gilovich (2001) later introduced the analogy of TOTE-units

from cybernetics (pp. 392-393) suggesting that what occurs when people adjust is a finite loop of

testing (Is this number a good estimate?), if no, *operating* (Take a bigger/smaller number),

testing again, and *exiting* (Take this number as the final estimate) when the testing is positive. This mechanism requires effort (e.g., Simmons et al., 2010, p. 917). We suggest that taking the loop multiple times or undertaking big adjustments is effortful. In terms of the TOTE-model, depleted participants would make *fewer steps* or *smaller* steps away from the anchor and therefore produce a final judgment closer to the anchor, as compared to non-depleted participants.

Finally, (5) scale distortion theory suggests "a change in the way numeric labels are being used in a particular context unaccompanied by any change in the beliefs or mental representation of the entity being judged" (Frederick & Mochon, 2012, p. 131). Note that Frederick and Mochon (2012, p. 131) "expect that scale distortion would have a negligible effect for most of the items [with self-generated anchors] (Epley & Gilovich, 2001, 2004, 2005, 2006), [as the mechanism] is restricted to judgments in which a strongly associated (and clearly relevant) referent is spontaneously brought to mind" (Frederick & Mochon, 2012, p. 131). We see no reason why ego depletion should have any effects on anchors if only this mechanism were at work. This means, if we find that depletion has no effect on anchoring this would support the notion that anchoring is based on scale distortion. If there is an effect, the other mechanisms are supported.

Hypotheses

Considering the different mechanisms proposed to underlie anchoring effects we argue that – except for scale distortion theory – all discussed theories and mechanisms should work for both self-generated and experimenter-provided anchors and derive the following hypothesis: Ego depletion leads to stronger anchoring effects, regardless whether the anchors were self-generated or provided by the experimenter.

This conforms with the *sticky anchor hypothesis of ego depletion* as proposed by Banker et al. (2017) but has not yet been tested with respect to internally generated anchors. In the case of null results (no main effect of depletion on the strength of anchoring effects), we conclude that the impact of ego depletion on the susceptibility to external cues as outlined by Banker et al. (2017) may depend on the social mechanisms that were present in their experiments and not on the internal adjustment and priming processes that are involved in anchoring effects. This would conform with results obtained by Haan and van Veldhuizen (2015). In the case of an opposite relationship, that is ego depletion leading to lower susceptibility to anchoring effects, we would conjecture that participants were less motivated to the degree that they gave random estimates or that participants became better at estimating due to unknown reasons. Such a finding would contradict the findings of Banker et al. (2017) and Epley and Gilovich (2001, Studies 2 and 3). If the sticky anchor hypothesis holds for external stimuli only (if ego depletion increases the strength of anchoring effects only for externally provided anchors), its universality should be restricted. Possible explanations could be different processing styles of internal and external stimuli.

In an exploratory fashion we will check for dose effects. The wide variability in effect sizes in studies of ego depletion suggests that, like any sort of fatigue and many other independent variables, there are varying degrees of it. However, relatively few studies have compared mild vs. severe depletion. One investigation by Vohs, Baumeister, and Schmeichel (2012) found that a stronger depletion effect produced not only larger effects but also some that were qualitatively different in terms of susceptibility to moderator variables. Anchoring and adjustment seems a potentially useful method for comparing different degrees of depletion, insofar as there are degrees of adjustment. We therefore included one measure after the first part

of the depletion essay and other measures after a more extensive manipulation (three essays). If there are significant differences between each pair of conditions (no depletion, mild depletion, and extensive depletion), that would be strong evidence for a dose-response relationship. If there is no effect after the mild manipulation but a significant effect after the extensive manipulation, that would extend the Vohs et al. (2012) conclusion that effects of severe depletion are qualitatively different from mild depletion.

Method

To test our hypothesis, we will test people's susceptibility to self-generated and experimenter-provided anchors that result in a clear direction of adjustment before and after an ego depletion procedure. All of the theory and methods parts of this article will be pre-registered and not altered after the data collection. As suggested by Simmons, Nelson, and Simonsohn (2012) we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Materials

Ego depletion method. We will use the "verbal flexibility task" from Sjåstad and Baumeister (2018) as it has been shown to yield a very strong effect on the manipulation check ($d = 4.71$), it can be easily standardized, and enables data to be collected via an online study. We will conduct an online study in order to achieve a large sample size which we will show to be necessary for this endeavor. In the task, participants will have to write three essays while not using certain letters. The essays' topics are a description of the neighborhood in which the participants grew up as a child (without the letters R or O), what participants do on a typical weekday (without the letters S or I), and a recent trip they had taken (without the letters A or N). There will be no control group as we will use a within-subject design. As in Sjåstad and

Baumeister (2018), participants will spend eight minutes for each of the essays which was determined by the questionnaire software. As we use a within-subject design, no manipulation check will be conducted. We will manipulate ego depletion within-subject, that is, we will measure susceptibility to anchoring effects before and after the verbal flexibility task. Note that besides the depletion model we have neither theoretical reason nor any evidence that suggests that anchoring effects get stronger or weaker over time.

Susceptibility to anchoring effects measure. To measure how strongly participants are influenced by anchors we created new items and adopted and translated items from past research (e.g., Chaxel, 2014; Simmons et al., 2010; Epley & Gilovich, 2001). A complete list of all items and anchors can be found online (osf.io/2c75d/). The anchors will consist of two kinds. A) Self-generated anchors include items adapted from Epley and Gilovich (2001, 2006), from Simmons et al. (2010), and new items. They are characterized by a prominent number "popping" into one's head. For example, when people are asked about the freezing point of vodka, they know that it is *below* (direction of adjustment) *0°C* (self-generated anchor) but they do not know the exact answer. We will check whether people actually knew the anchor or used it for their estimation. If not, their answers will be excluded on an item-by-item basis as in previous research (e.g., Epley & Gilovich, 2006, p. 314). To ensure a minimal number of exclusions through this procedure we conducted a pretest to determine the adequacy of 13 different items. Based on the proportion of people who both knew the intended anchor and thought of it when estimating we excluded four items. B) Experimenter-provided anchors will include items adopted and translated from Simmons et al. (2010) and new items. For each item, the direction of adjustment will be certain by giving hints, such as "the correct value is higher than 10".

Further measures. For exploratory purposes we will measure participants' self-control. Therefore we will use a German translation of the Brief Self-Control Scale (Sproesser, Strohbach, Schupp, & Renner, 2011).

To test for dose effects we will add one self-generated anchor item and one experimenter provided anchor item before and after the first essay to explore whether administering one essay only to participants has the same effect as three essays have. A list with the exploratory items can be found in the supplementary materials online (osf.io/2c75d).

Procedure

After being welcomed to the online study, participants will be told that the study is about general knowledge questions and writing essays. Then they answer demographic questions (sex, age). Ten questions for measuring anchoring effects will be randomly drawn without replacement of which five are from the pool of self-generated anchor items and five are from the pool of experimenter-provided anchors. Two of these ten items come from a different pool and are used for exploratory purposes only. These ten questions are followed by the ego depletion task. Between essays 1 and 2, two more exploratory anchor items are administered. After participants have finished all essays, they will have to answer eight questions for measuring anchoring effects. They will be presented in the manner described above. For self-generated anchor questions we will check for each participant whether he or she a) knew the anchor and b) thought of the anchor value when generating the estimate (e.g., Epley & Gilovich, 2001) in order to assess whether the anchor has actually been generated by the participants. At the end of the experiment, participants will complete the Brief Self-Control Scale and receive feedback about their estimates and the correct answers to the anchoring questions. The estimated duration of the

experiment is approximately 35 minutes. The questionnaire file (German) can be accessed online (osf.io/2c75d). The procedure can be seen in Figure 1.
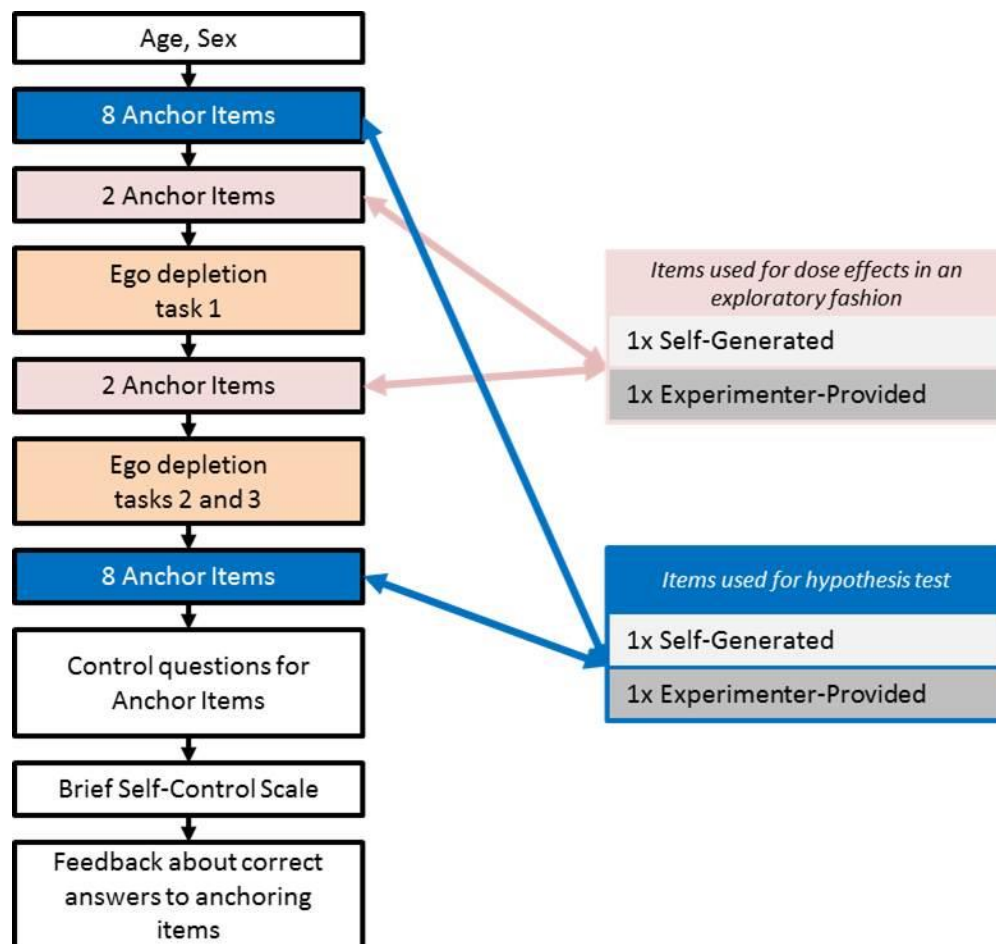
Figure 1: Procedure of the study.

Planned sample

Data will be collected online via social networks and university newsletters. Participants will be rewarded with feedback of their performance during the anchoring tasks and the chance to win one of two 20€ Amazon vouchers. A sample feedback letter can be found online (osf.io/2c75d).

To determine the sample size we compared effect sizes of our planned ego depletion procedure and of several variables that influence the susceptibility to self-generated anchors (see

Table 1). The experiment by Haan and van Veldhuizen (2015) in which no effect of ego depletion on anchoring could be found is not included in the table. We decided to use the minimum effect size $d = .15$ and thus will need 483 participants to achieve 95% power ($\alpha = .05$, one-tailed t-test for dependent groups, equal sample sizes). We determined to stop data acquisition once 500 people have participated successfully in the experiment. "Participated successfully" means, that they were not excluded due to reasons specified below. With this big a sample, interaction effects of the depletion state (before vs. after depletion procedure) with the kind of anchor (self-generated vs. externally provided) of $f = .110$, which is considered a small to medium effect, can be found. Cases will be excluded if any of the following criteria is met: missing values on any of the variables that are required for the hypothesis tests (e.g., through dropouts), overly consistent responses (i.e. standard deviation of the anchoring questions is 0), age below 18. Participants will also be excluded if they write very little (i.e. less than 50 characters in any of the essays of the depletion manipulation). Criteria for the susceptibility to anchoring effects measure apply on an item-by-item basis and are the following: estimates in the anchoring tasks that are imprecise (e.g., intervals), equal to the provided anchor value, outliers (defined as ±3SD). Moreover, participants' answers to questions with self-generated anchors will be excluded if they answer no to any of these two questions (adapted from Epley & Gilovich, 2006, p. 314): 1) Do you know the anchor? 2) To give your estimate, did you think of that value? Finally, participants who do not have at least one estimate per anchor type (self-generated vs. experimenter-provided) and per time (before vs. after ego depletion) are excluded.

Table 1: Effect sizes of ego depletion on variables and of variables on self-generated anchors.

Study	IV	DV	*d*	*n*
Sjåstad & Baumeister, 2018	Ego depletion	Willingness to plan	.62	105
Sjåstad & Baumeister, 2018	Ego depletion	Manipulation Check	4.71	105
Epley & Gilovich, 2006, Study 2a, p. 316	Need for Cognition	Self-generated anchors	.49	79
Epley & Gilovich, 2006, Study 2b, p. 316	Alcohol consumption	Self-generated anchors	.85	52
Epley & Gilovich, 2006, Study 2c, p. 316	Cognitive load	Self-generated anchors	.66	94
Epley & Gilovich, 2001, Study 2, p. 394	Head movement	Self-generated anchors	.36	93
Epley & Gilovich, 2001, Study 3, p. 395	Head movement	Self-generated anchors	.50	32
Epley & Gilovich, 2005, Study 1, p. 205	Monetary incentive	Self-generated anchors	.52	56
Epley & Gilovich, 2005, Study 2, p. 208	Forewarning	Self-generated anchors	.51	48
Epley & Gilovich, 2004, Study 1, p. 450	Arm movement	Self-generated anchors	.56	34
Simmons et al., 2010, Study 2, p. 923-924	Monetary incentive	Self-generated anchors, direction of adjustment is known	.21	125
Simmons et al., 2010, Study 5, p. 929	Motivation	Self-generated anchors, direction of adjustment is known	.23	115
Banker, et al. 2017, Study 1, p. 3	Ego depletion	Taken money in a reversed dictator game	.66	28
Banker, et al. 2017, Study 2, p. 5	Ego depletion	Taken money in a reversed dictator game	.31	376
Banker, et al. 2017, Study 3, p. 7, low anchor condition	Ego depletion	Taken money in a reversed dictator game	.22	~741/2
Banker, et al. 2017, Study 3, p. 7, high anchor condition	Ego depletion	Taken money in a reversed dictator game	.15	~741/2

Data Analysis Plan

The procedure to process estimates on the anchoring questions were adopted from

Simmons et al. (2010, p. 923). We will compute absolute differences between the anchor value

and the estimate. If multiple anchor values are allowed (e.g., the human body temperature can be estimated to lie between 36-38°C) the difference between the estimate (e.g., what is the lowest recorded body temperature in a human being?) and the self-generated anchor value as indicated in the control question (e.g., what is the average human body temperature?) will be computed. We z-score these values separately for each question and independently of which experimental group the item belonged to. We averaged for each participant the question's z-score value by anchor type (self-generated vs. experimenter-provided) for the four items that precede and proceed the ego depletion procedure separately. When interpreting values, higher positive numbers indicate a larger gap between anchors and estimates. Henceforth we refer to this score as participants' *adjustment-scores*. Each participant's adjustment-scores will follow a 2 (self-generated vs. experimenter-provided) by 2 (before depletion task vs. after depletion task) design.

In order to test our hypothesis that ego depletion leads to stronger anchoring effects, we will conduct a one-tailed t-test for dependent groups with before vs. after the ego depletion procedure (i.e. the verbal flexibility task). To test the interaction of the strength of the anchoring effect with whether the anchor was self-generated or not, we compute a 2 (self-generated vs. experimenter-provided anchor value) by 2 (before depletion task vs. after depletion task) mixed ANOVA. The dependent variables will be participants' average adjustment-scores. The analysis script we will use can be found online (osf.io/2c75d/). Data analysis will be conducted with R (R Core Team, 2018) using the packages Reshape (Wickham, 2007) and Stringr (Wickham, 2018).

References

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The Sticky
Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues. *Journal of Behavioral Decision Making*, *87*(1), 23. https://doi.org/10.1002/bdm.2022

Baumeister, R. F., Tice, D. M., & Vohs, K. D. (2018). The Strength Model of Self-Regulation:
Conclusions From the Second Decade of Willpower Research. *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, *13*(2), 141–145.
https://doi.org/10.1177/1745691617716946

Bertrams, A., Baumeister, R. F., Englert, C., & Furley, P. (2015). Ego depletion in color priming
research: Self-control strength moderates the detrimental effect of red on cognitive test
performance. *Personality & Social Psychology Bulletin*, *41*(3), 311–322.
https://doi.org/10.1177/0146167214564968

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

Dang, J., Liu, Y., Liu, X., & Mao, L. (2017). The Ego Could Be Depleted, Providing Initial
Exertion Is Depleting. *Social Psychology*, *48*(4), 242–245. https://doi.org/10.1027/1864-9335/a000308

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. (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

Fischer, P., Greitemeyer, T., & Frey, D. (2008). Self-regulation and selective exposure: The impact of depleted self-regulation resources on confirmatory information processing. *Journal of Personality and Social Psychology*, *94*(3), 382–395. https://doi.org/10.1037/0022-3514.94.3.382

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology. General*, *141*(1), 124–133. https://doi.org/10.1037/a0024006

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2018). Is Ego Depletion Real? An Analysis of Arguments. *Personality and Social Psychology Review : an Official Journal of the Society for Personality and Social Psychology, Inc*, 1088868318762183. https://doi.org/10.1177/1088868318762183

Haan, T. de, & van Veldhuizen, R. (2015). Willpower depletion and framing effects. *Journal of Economic Behavior & Organization*, *117*, 47–61. https://doi.org/10.1016/j.jebo.2015.06.002

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495–525. https://doi.org/10.1037/a0019486

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality & Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Mussweiler, T., & Strack, F. (1999). 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., & Pfeiffer, T. (2000). Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142–1150. https://doi.org/10.1177/01461672002611010

R Core Team. (2018). R. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*(1), 33–46. https://doi.org/10.1037/0022-3514.85.1.33

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2160588

Sjåstad, H., & Baumeister, R. F. (2018). The Future and the Will: Planning requires self-control, and ego depletion leads to planning aversion. *Journal of Experimental Social Psychology*, *76*, 127–141. https://doi.org/10.1016/j.jesp.2018.01.005

Sproesser, G., Strohbach, S., Schupp, H., & Renner, B. (2011). Candy or apple? How self-control resources and motives impact dietary healthiness in women. *Appetite*, *56*(3), 784–787. https://doi.org/10.1016/j.appet.2011.01.028

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

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science*, *11*(3), 249–254. https://doi.org/10.1111/1467-9280.00250

Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology*, *48*(4), 943–947. https://doi.org/10.1016/j.jesp.2012.03.002

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, *21*(12). https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2018). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.0. Retrieved from https://CRAN.R-project.org/package=stringr

Article 2: Subliminal Anchoring

The following manuscript is the submitted version of a manuscript that is currently under review at the Journal of Experimental Social Psychology.

Röseler, L., Schütz, A., Blank, P. A., Dück, M., Fels, S., Kupfer, J., Scheelje, L., & Seida, C. (2020). No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered, and Failed Replication Attempts. *Manuscript submitted for publication*.

No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered, and Failed Replication Attempts

Lukas Röseler [1,2], Astrid Schütz[1], Pia A. Blank[1], Marieluisa Dück[1], Sabine Fels[1], Jana Kupfer[1],

Linda Scheelje[1], and Christian Seida[1]

[1]University of Bamberg, [2]Harz University of Applied Sciences

Abstract

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

Keywords: subliminal, anchoring, anchoring effect, numeric priming, replication

No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered, and Failed

Replication Attempts

When numeric estimates are biased toward a previously considered numeric value, an

anchoring effect has occurred. This phenomenon has received remarkable attention ever since

Tversky and Kahneman (1974) described it. Numerous theories have been invented (e.g.,

Frederick & Mochon, 2012; Mussweiler & Strack, 1999a), claims about how this is one of the

most important truths have been made (Simmons, LeBoeuf, & Nelson, 2010, p. 917), many

moderators have been identified (for an overview, see Furnham & Boo, 2011), and classical

effects have been successfully replicated (Klein et al., 2014).

Recently, however, a tidying-up has begun in the anchoring literature, and a graveyard

has been created for deceased theories (Ferguson & Heene, 2012). Researchers have started to

criticize the lack of empirical content in current theories and are now going back to the findings

that sparked these theories only to find that (a) some of these findings cannot be replicated

(Harris et al., 2019; Klein et al., 2018), (b) the theories are wrong (e.g., the claim that the

insufficient adjustment model is valid for only self-generated anchors and the selective

accessibility model is valid for only experimenter-provided anchors has been refuted; Chaxel,

2014; Epley & Gilovich, 2001; Simmons et al., 2010), and (c) the findings were questionable in

the first place (e.g., numerous traits have been identified or suspected to moderate the

susceptibility to anchoring even though the susceptibility to anchoring cannot yet be measured

reliably; Cheek & Norem, 2019; Röseler, Schütz, & Starker, 2019). Replication is finally starting

to gain acceptance in psychology and is important for both its progress and the public's

perception of it (Wingen, Berkessel, & Englich, 2019). We acknowledge both productive and

destructive efforts as vital parts of psychological science. With this report, we join the effort to resolve a conflict about anchoring that has existed for 14 years in the literature and—to our knowledge—has yet to receive any attention.

Subliminal Anchoring

To our knowledge, there are only two published studies that have investigated *subliminal anchoring*. Subliminal anchoring is a special case of anchoring in which the anchor is not actually explicitly mentioned as in other paradigms that often feature the comparative question "Is it more or less than X?" Thus, subliminal anchoring has a great deal of overlap with incidental anchoring and priming. In incidental anchoring, the anchor is explicitly presented to people (e.g., when the number in a cell phone's name is P17 vs. P97), but there is no comparative question. Consecutive numeric estimates (e.g., the proportion of items sold) have still been found to be biased toward the number given in incidental priming (Critcher & Gilovich, 2008). However, the effect was not replicated (Edmonds; Klein et al., 2018). Likewise, priming has faced replication issues (e.g., Cesario, 2014). In subliminal anchoring, there is no comparative question, and the anchor is presented in the participants' field of view for a very short time so that participants do not notice the number. The two published studies on subliminal anchoring presented anchors for 33 ms (Mussweiler & Englich, 2005) or 17 ms (Reitsma-van Rooijen & Daamen, 2006).

Contradictions between Subliminal Anchoring Studies

The two existing studies both found subliminal anchoring effects. However, the conditions that have been proposed to be necessary for subliminal anchoring are contradictory. Mussweiler and Englich (2005) provided the first published evidence of subliminal anchoring. As part of their cover story, they asked participants to concentrate and *take their time* arriving at

an estimate. During this period of concentration, the anchor was presented. By contrast, Reitsma-van Rooijen and Daamen (2006) reported that subliminal anchoring works only when participants are under time pressure, which is the opposite of what Mussweiler and Englich suggested.

Neither paper provided a thorough theoretical account for why (no) time pressure should be essential or how subliminal anchors actually work. Of course, the selective accessibility model of anchoring (e.g., Mussweiler & Strack, 1999b) could be applied to subliminal anchoring because it uses priming as one of its mechanisms. However, given the recent failure to replicate one of its key findings and its contradictory predictions in some domains, it does not seem to provide a clear theoretical account of subliminal anchoring. For example, ego depletion is predicted either to increase (Banker, Ainsworth, Baumeister, Ariely, & Vohs, 2017) or to decrease (Francis, Milyavskaya, Lin, & Inzlicht, 2018) the strength of anchoring on the basis of the selective accessibility model, but there was no effect at all in a registered report (Röseler, Schütz, Baumeister, & Starker, 2020).

With the present paper, we are attempting to clarify the contradiction with respect to time pressure between the two contradictory studies. To do so, we conducted high-powered preregistered replications of the two studies. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012).

Study 1: Replication of Mussweiler and Englich (2005), Study 2

Mussweiler and Englich (2005) demonstrated anchoring effects when participants concentrated on a computer screen for 1 min and anchors were subliminally presented 10 times for 33 ms each time and masked with symbols (e.g., $§?#ß#). High and low anchors influenced

the estimate of the mean temperature in Germany (Study 1) and the price of a middle-class car (Study 2). Apart from anchors and the target question, the only other difference between Mussweiler and Englich's Studies 1 and 2 was the prime detection task in Study 2, in which participants tried to recognize the subliminally presented anchor. Study 2 showed that participants could not recognize subliminal anchors even when participants knew the anchors existed and wanted to recognize them.

Method

As in the original study, in the current study, participants were presented with a question and were then asked to focus on flickering letters (among which the anchor was presented) and subsequently to provide an estimate regarding the target question. Afterwards, participants completed the prime detection task: They were told about the subliminal anchors and were asked to guess the numbers presented in 10 trials. The result was used to test whether their performance was better than chance and whether the presentation was actually subliminal. The presented numbers were randomized such that the high anchor was presented five times, and the low anchor was presented five times. We used the funneled debriefing method to test for participants' awareness of the primes, which included seven awareness check questions and which revealed the purpose of the experiment (Mussweiler & Englich, 2005, p. 136). In a personal correspondence, the first author of the original study revealed that the original materials were no longer available (the original study was conducted more than 14 years ago). We thus designed our study to resemble the original study as closely as possible using the details described in the original publication. We preregistered the replication before data collection (https://www.osf.io/8urxv) using the replication recipe (Brandt et al., 2014; the postcompletion registration is available at https://www.osf.io/acykd).

Planned sample size. To facilitate a power analysis, we pooled the effect sizes of the original Studies 1 and 2 (which differed only with respect to the target question) and found a mean effect of $d = 0.69$ ($N = 76$, 95% CI [0.23, 1.16], $N_{min95\%} = 94$) of subliminal anchoring. The present study was conducted as part of a research class and thus the target sample size was also chosen for practical reasons. We decided to collect data from $N = 160$ participants, which would yield $1 - \beta > .99$ for $d = 0.69$ (original effect size) and $1 - \beta > .80$ for $d = 0.40$ (mean effect size for a variety of anchoring effects).

Deviations from the original study. Our study deviated from the original study in several ways. (a) We could not conduct the study in a lab, but we instructed participants to do the experiment in a calm working environment. (b) To achieve our target sample size, we recruited not only students but also nonstudents. (c) We did not reward participants as had been done in the original study by giving participants 6 € (Study 1) or an ice cream cone (Study 2) after the study. However, those participants had probably not been told about the ice cream beforehand, and the provision of the cone was not linked to performance. Thus, not providing a reward can be viewed as only a minor deviation in terms of replication closeness. (d) As in the original study, we conducted a small pretest to determine the high and low anchor values. The mean estimate of the price of a middle class car was $M = 23,026.32$ € ($SD = 6,441$, $N = 19$), which is why we chose 10,000 and 40,000 as anchors. (e) We decided to document whether the participants were students to see if there was a difference between students (i.e., the participants in the original study) and nonstudents.

Exclusion criteria. In Mussweiler and Englich's (2005) Studies 1 and 2, the published exclusion criteria were ambiguous. They gave two reasons for the exclusions. (a) A participant who recognized all 10 subliminal anchors in the prime detection task was removed from the

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

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

Results

Achieved sample size. A total of 183 participants completed the experiment (see replication_rawdata.csv file at https://osf.io/96de2/files/). Six participants were excluded because they reported all 10 subliminal anchors correctly; 64 participants would have been excluded

because they noticed the flickering on the screen (for a discussion, see the Method section); 12 participants were excluded because they said they had recognized the subliminal anchors. Of the remaining 165 participants, 80 had been presented with the high and 85 with the low anchor; 93 were women, 67 men, and 3 diverse; 136 were students. The mean age was $M = 24.12$ years ($SD = 6.95$ years). The achieved power for $d = 0.69$ (pooled original effect size) was thus $1 - \beta > .999$.

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

Exploratory analyses. All further analyses were not preregistered and were conducted for exploratory purposes only.

Excluding outliers. There were five outliers in both the high and the low anchor group ($Max = 300,000$, $Z_{max} = 8.19$). In an analysis that was not preregistered, we applied an exclusion criterion that is common in anchoring research, that is, we excluded values that were three standard deviations above or below the mean estimate (e.g., Chaxel, 2014, p. 47). The estimates in the high anchor condition ($M = 20,763.89$, $SD = 11,136.08$, $N = 65$) were not significantly

larger than the estimates in the low anchor condition ($M = 21,100.00$, $SD = 9,949.01$, $N = 71$), $t(128.82) = -0.18$, $p = .573$ (one-tailed), $d = -0.032$.

Susceptibility to anchoring and prime detection task. Mussweiler and Englich (2005) tested for a correlation between the susceptibility to the subliminal anchor (i.e., absolute difference between the estimate and anchor) and the number of correctly recognized primes in the prime detection task. As in the original study, we did not find a correlation, $r(134) = .049$, $p = .572$ (two-tailed). This result was not sensitive to including or excluding the outliers.

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

Participant age. By including nonstudents, our sample's mean age was most likely higher than the mean age from the original study (but Mussweiler & Englich, 2005, reported no demographic data). It is possible that the optimal calibration of the subliminal anchor's presentation time further depends on participants' age. We thus tested for whether age was correlated with the number of correctly recognized subliminal anchors in the prime detection task, but we found a very small and nonsignificant correlation, $r(161) = -.152$, $p = .053$ (two-tailed).

Discussion

Our high-powered, preregistered, close replication of Mussweiler and Englich (2005) may have been miscalibrated because the anchors were not necessarily presented subliminally.

However, as subliminal anchoring is a special case of numeric priming, and as this priming should be more likely to occur when participants perceive the numeric value consciously, the effect would have been easier to replicate if the anchors had not been subliminal. However, there was no effect, and mean estimates in the high anchor condition were even lower than in the low anchor condition. Due to ambiguous reports in the original study, we changed one preregistered exclusion criterion but found that whether or not the participants recognized the flickering had no impact on the strength of the anchoring.

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

Study 2: Replication of Reitsma-van Rooijen and Daamen (2006)

Our results are not the first to call into question Mussweiler and Englich's (2005) claim about the existence of subliminal anchors. One year after Mussweiler and Englich's study came out, Reitsma-van Rooijen and Daamen (2006) published results of an experiment on subliminal anchoring and reported that subliminal anchoring occurred only under time pressure. In their experiment, which was slightly different from the one by Mussweiler and Englich, participants were first exposed to the subliminal anchors (10 vs. 90) for only 17 ms (instead of 34 ms in the study by Mussweiler & Englich, 2005) and then asked to provide an estimate of the probability of the recurrence of a pestilence of the lungs in India within a year. Participants were asked/not asked to provide their answer quickly. Anchoring occurred only in the condition in which participants had to give their answer quickly. The null findings in Mussweiler and Englich's

Study 1, in which participants were asked to take their time, could thus be explained by Reitsma-van Rooijen and Daamen's explanation. We conducted a high-powered, preregistered, close replication of the study by Reitsma-van Roijen and Daamen.

Method

In the present study, just as in the study by Reitsma-van Rooijen and Daamen (2006), participants were greeted and then asked to complete demographic questions (gender, age, and whether they were students). Then they were told that this study was about how people make estimates and that "we knew from experience that participants were able to give the correct answer, when they followed their first impression" (Reitsma-van Rooijen & Daamen, 2006, p. 383). They were introduced to the letter task, which required them to say whether a random letter combination consisted of more lower or more upper case letters. One difference was that participants who had completed the study in our lab were given candy. Furthermore, we preregistered the replication before data collection (https://www.osf.io/qxwhv) using the replication recipe (Brandt et al., 2014; the postcompletion registration is available at https://osf.io/gh52k).

Letter task. In the original study, there was an exercise to familiarize participants with the letter task. As this exercise was not described and we were unable to obtain information from the authors, we created a new one using the description in the publication: Participants completed five trials of the very same task that they would encounter later, except that there were no subliminal anchors. In the letter task, (a) two crosses ("XX") were presented for 500 ms in the center of the screen, (b) then the anchor (low anchor: 10; high anchor: 90) was presented for 17 ms (this was omitted in the exercise), (c) the two crosses were presented for 17 ms, (d) the letter combination was presented for 1,500 ms, and then (e) the question appeared "Were there more

upper or lower case letters?" until participants gave their answer by pressing either A (upper) or L (lower) on the keyboard. After 15 trials, participants saw the message "This was the last letter combination; please make the probability estimate [please give your answer quickly]" for 5,000 ms. Afterwards, participants typed their estimate. The request for a quick answer was omitted from the no-time-pressure condition.

Letter task stimuli. The original study used Chicago font, plain, 14 pt., 4.5 mm high from a 0.5 m viewing distance. Because Chicago font is no longer used on contemporary computers, we used a similar sans serif font that is common across most computers, namely, Tahoma. We made sure that letters were 4.5 mm high and were viewed from a 0.5 m viewing distance in the lab. We also excluded upper case and lower case letters that were likely to be confused, such as c/C, k/K, and v/V. A list of all the letters we used can be found online (letter_task.xlsx at https://www.osf.io/96de2/files/). In the original study, there were 15 letter combinations, and "there were as many capital letters as lower case letters added over the letter combinations" (Reitsma-van Rooijen & Daamen, 2006, p. 383). Due to a lack of further information, we decided that all letter combinations would have a 4:2 or 2:4 ratio of upper:lower case letters. Using pseudo-random numbers, we generated 8 combinations with a 4:2 ratio and 8 combinations with a 2:4 ratio. Each participant saw 15 of these 16 combinations. For the exercise, we created 2 times 3 combinations using the same ratios.

Planned sample size. The original effect sizes were $d = -0.284$ for the no-time-pressure condition and $d = 0.867$ for the time-pressure condition. A total of 62 participants (15 or 16 participants per cell) completed the original study. As the authors of the original study found that there was no effect in the no-time-pressure condition, we predicted an interaction effect size of f $= (d1-d2)/4 = (0.867-0)/4 = 0.21675$. Because achieving 95% power ($N_{min} = 279$) was not

possible for practical reasons, we aimed for 80% power and thus a target sample size of $N = 170$, which was still almost three times the original sample size.

Deviations from the original study. We deviated from the original study in several ways. (a) We did not present the question for 5,000 ms but for as long as participants needed to type in their estimate and click on the "next" button ($Mdn_{response\ time} = 18.36$ s). (b) We designed a new exercise and new stimuli for the letter task. (c) We asked for demographic data (i.e., gender, age, and student status). (d) We applied preregistered exclusion criteria. (e) Finally, due to the SARS-CoV-2 pandemic, we collected some of our data online rather than in the lab.

Exclusion criteria. As percentages have to be between 0 and 100, all participants who estimated a number below 0 or above 100 were excluded. Second, if participants guessed the answer and did not concentrate on the letter combinations, they missed the subliminal anchor. This is why we excluded participants if they gave correct answers for only three or fewer of the 15 letter task trials. Trials from the exercise were not considered.

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

Results

Achieved sample size. Due to the Coronavirus pandemic, which changed everyday life around the world, we had to stop recruitment in our laboratory and continue the survey online (thereby deviating from the preregistration). A total of 111 participants were administered the original version in our labs, whereas 59 participants took the online version; 175 participants completed the experiment (see replication_rd_rawdata.csv file at https://osf.io/96de2/files/). No participant was excluded because of estimated probabilities larger than 100%. No participant was

excluded due to too few correct responses in the letter task. Three participants were excluded because they recognized that the experiment was about anchoring or subliminal perception. The remaining sample was reduced to the target sample size of $N = 170$, and 80% power was thus achieved (see analysis script and replication_rd_processed.csv file at https://osf.io/96de2/files/). Participants consisted of 121 women, 48 men, and 1 diverse; 114 were students. Their mean age was $M = 31.97$ years ($SD = 17.19$ years).

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

Hypothesis test. A 2 (time pressure: no vs. yes) × 2 (anchor: low vs. high) ANOVA revealed no main effect of the anchor, $F(1, 166) = 1.04$, $p = .310$, $f = 0.079$. Time pressure did not have an effect on the estimate, either, $F(1, 166) = 0.06$, $p = .810$, $f = 0.019$. Whereas we hypothesized a Time Pressure × Anchor interaction, no such effect was present, $F(1, 166) = 0.11$, $p = .743$, $f = 0.025$, 95% CI [0, 0.155]. This interaction effect was significantly smaller than the original effect of $f = 0.867$. Contrary to the original effect, post hoc analyses revealed that the anchoring effect was descriptively stronger in the no-time-pressure condition ($d_{no\ time\ pressure} = 0.226$ vs. $d_{time\ pressure} = 0.104$). The data for the hypothesis test are presented in Table 1 and *Figure 1.*

Exploratory analyses. All further analyses were not preregistered and were computed for exploratory purposes only.

We checked whether participants in the online version had lower concentration by comparing the scores in the letter task between the data from the lab and the online data. The mean letter task score was higher for the participants in the lab (M_{lab} = 13.61, SD_{lab} = 1.40, N_{lab} = 111) than for the online participants (M_{online} = 13.52, SD_{online} = 1.30, N_{online} = 59), but the difference was not significant, $t(126.06) = 0.40$, $p = .687$ (two-tailed), $d = 0.061$.

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

To test whether the letter task influenced the strength of the anchoring effect, we correlated the absolute difference between the estimate and the anchor with the letter task score. The correlation was almost zero, $r(168) = .015$, $p = .845$ (two-tailed).

Table 1

Descriptive Statistics for the Probability Estimate used in the Hypothesis Test

Anchor	Time pressure	M	SD	N
Low	Absent	17.82	17.82	44
High	Absent	22.17	21.12	35
Low	Present	17.96	20.92	53
High	Present	20.21	22.60	38
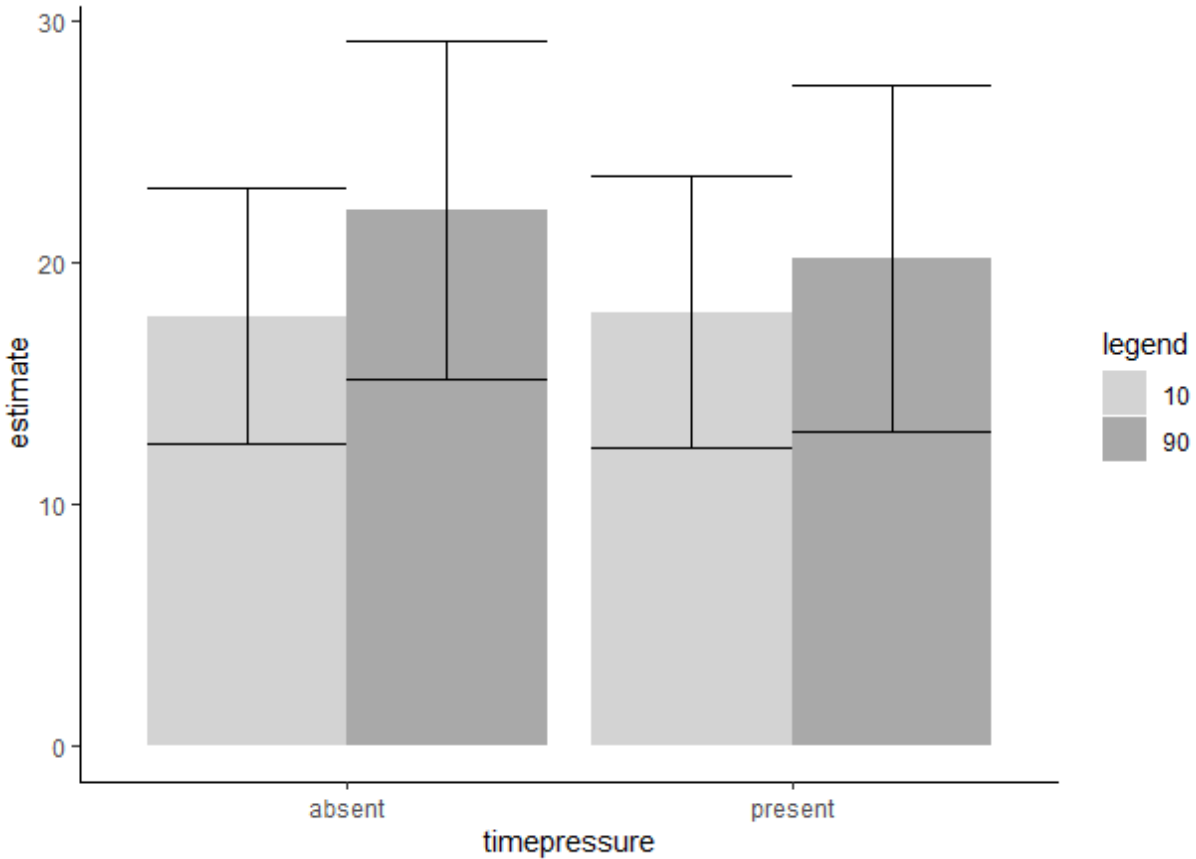
Figure 1. **Nonsignificant interaction between time pressure and subliminal anchor for predicting the estimates.**

Discussion

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

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

The more severe issue might be that we deviated from our preregistration by ending recruitment in our lab and switching to an online version of the experiment. This put the (lack of) incentive back in line with the original study (in the original study, no incentive was reported, whereas participants in our lab were given candy). Because the experiment was hosted online anyway (and the participants in the lab sat in front of a computer screen), none of the code had to be changed. For the online sample, we were unable to check for whether the computer screen's refresh rates were at least 60 Hz, nor could we cover any numbers presented in the rooms where participants took the experiment in their homes. Moreover, online recruitment meant that nonstudents could take part in the experiment. In the original study, only students from Leiden University participated in the experiment. However, as our sample size was approximately three times the original sample size, and as the sample size of the online version was as large as the original study, we computed exploratory analyses to determine that the location of the experiment and the student or nonstudent status did not matter for the hypothesis test.

General Discussion

In two high-powered, preregistered, close replications, we found no evidence for subliminal anchoring. The quality of the data was determined to be high and, therefore, cannot

provide an explanation for the failed replication attempts. We have openly addressed all

deviations from the original studies and our preregistration and conducted exploratory analyses

to ensure that they had no relevant impact. Results from the three original studies and our

replications are presented in Table 2.

Table 2

Sample Sizes and Cohen's ds for the Subliminal Anchoring Effect in the Original and Replication Studies

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

Limitations

The most severe limitations of our replication studies are the deviations from the original

studies. Although our replications were as close as possible to the original studies, this does not

mean that there were no deviations. While designing our studies, we noted that Reitsma-van

Rooijen and Daamen (2006) reported many more details than Mussweiler and Englich (2005).

Closeness could thus be established more easily for the second replication study. Note moreover

that the original materials were no longer available. Guaranteeing closeness was furthermore

impeded by ambiguous or missing descriptions. For example, Mussweiler and Englich (2005)

did not clarify what they meant when they stated that participants "indicated their suspicion

about the flickering of the letter string," nor did they identify the size of the letters on the screen.

Reitsma-van Rooijen and Daamen (2006), on the other hand, had not described what exercise

they used to familiarize participants with the letter task. In both cases, the closeness between our replications and the original studies is impossible to determine.

In a recent discussion about replicator's degrees of freedom (Yeager, Bryan, & O'Brien, 2019), some replicators have been accused of exploiting researcher degrees of freedom (Wicherts et al., 2016) to produce nonsignificant results. We used the replication recipe preregistration template[9] (Brandt et al., 2014) and sought to ensure that we did not exploit researcher degrees of freedom. This included methods of data analysis. The analytic choices had been made prior to data analysis, and the models were not complex (*t* tests and ANOVA). The analysis code for the experiments and R scripts were written before we conducted the experiments. All changes in the code are highlighted.

The Future of (Subliminal) Anchoring

Although it has been called one of the most robust phenomena (Simmons et al., 2010, p. 917), anchoring is no exception to the omnipresence of findings that are difficult to replicate. Instead of moving forward and building on brittle bricks, we resolved an old contradiction between two findings by showing that neither of them could be replicated. This raises the question of why there are no other publications on the topic and why the contradiction had not yet been resolved. There are several possibilities, none of which seems promising for the state of anchoring research: (a) nobody recognized the contradiction, (b) everybody knew that the

---

[9] That said, we encountered disadvantages of the replication recipe that other templates such as the social psychology template by van 't Veer and Giner-Sorolla (2016) solved effectively: There is no way to upload files such as study materials or analysis scripts in the replication recipe template. Adding this seems to be a small but effective step.

findings could not be replicated, (c) whether these findings could or could not be replicated is not important, or (d) there have been other attempts to show that subliminal anchoring does not exist, but the results could not be published.

Preregistration in anchoring research is rare but not absent (e.g., Harris et al., 2019; Klein et al., 2018). Thus, the robustness of anchoring may have been overestimated. To avoid problems such as vastly overestimated effect sizes that have been uncovered in other fields by large multilab preregistered replications (e.g., Dang et al., 2020; Hagger et al., 2016), which have uncovered years of underpowered studies, we join Inzlicht and Friese (2019) in their call for a moratorium on nonpreregistered experiments. By this, we do not aim to hamper pluralistic research but to allow for cumulative psychological science. This is not yet possible as the current public record in many fields merely reflects lucky researchers and smart p-hackers.

Open Practices

All materials including the, analysis scripts, questionnaire files, stimulus generation code, and the data sets are available online at https://www.osf.io/96de2/ (Study 1 Mussweiler & Englich, 2005, replication preregistration: https://www.osf.io/97gaf; postcompletion registration: https://www.osf.io/acykd; Study 2 Reitsma-van Rooijen & Daamen, 2006, replication preregistration: https://www.osf.io/qxwhv/; postcompletion registration: https://osf.io/gh52k). All analyses have either been preregistered or marked as exploratory. We highlighted all deviations from the preregistration and the analysis script. All preregistrations and all studies that we conducted on subliminal anchoring were reported. We encourage other researchers to use our data and materials or to conduct different replication studies to clarify the contexts in which subliminal anchoring can occur.

Contributions

- LR developed the research questions, drafted the analysis scripts, and programmed the studies.

- LR and AS oversaw the project and drafted the manuscript.

- All authors designed and conducted Study 1 and the respective preregistration.

- All authors revised and approved the final manuscript.

References

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The Sticky

    Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues. *Journal of*

    Behavioral Decision Making, *87*(1), 23. https://doi.org/10.1002/bdm.2022

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . van 't

    Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal*

    of Experimental Social Psychology, *50*, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Cesario, J. (2014). Priming, Replication, and the Hardest Science. *Perspectives on Psychological*

    Science : A Journal of the Association for Psychological Science, *9*(1), 40–48.

    https://doi.org/10.1177/1745691613513470

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

Cheek, N. N., & Norem, J. K. (2019). Are Big Five Traits and Facets Associated With

    Anchoring Susceptibility? *Social Psychological and Personality Science*, *92*,

    194855061983700. https://doi.org/10.1177/1948550619837001

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral*

    Decision Making, *21*(3), 241–251. https://doi.org/10.1002/bdm.586

Dang, J., Barker, P., Baumert, A., Bentvelzen, M., Berkman, E., Buchholz, N., . . .

    Zinkernagel, A. (2020). A Multilab Replication of the Ego Depletion Effect. *Social*

    Psychological and Personality Science, 194855061988770.

    https://doi.org/10.1177/1948550619887702

Edmonds, K. A. Subliminal anchoring effects on weight estimates: Numerals versus numerosities. Advance online publication. https://doi.org/10.13140/RG.2.2.17238.24649

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

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a Within-Subject, Repeated-Measures Ego-Depletion Paradigm. *Social Psychology*, *49*(5), 271–286. https://doi.org/10.1027/1864-9335/a000348

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology. General*, *141*(1), 124–133. https://doi.org/10.1037/a0024006

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., . . . Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., . . . Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory

of anchoring. *Journal of Experimental Psychology. General.* Advance online publication.

https://doi.org/10.1037/xge0000644

Inzlicht, M., & Friese, M. (2019). The Past, Present, and Future of Ego Depletion. *Social

Psychology*, *50*(5-6), 370–378. https://doi.org/10.1027/1864-9335/a000398

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J. [Michael

J.], . . . Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*,

45(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., . . .

Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Sample

and Setting. Advance online publication. https://doi.org/10.31234/osf.io/9654g

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and

underlying mechanisms. *Organizational Behavior and Human Decision Processes*, *98*(2),

133–143. https://doi.org/10.1016/j.obhdp.2004.12.002

Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of

judgemental anchoring. *European Review of Social Psychology*, *10*(1), 135–167.

https://doi.org/10.1080/14792779943000044

Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the

anchoring paradigm: A selective accessibility model. *Journal of Experimental Social

Psychology*, *35*, 136–164. https://doi.org/10.1037/e413782005-562

Reitsma-van Rooijen, M., & Daamen, D. D., L. (2006). Subliminal anchoring: The effects of

subliminally presented numbers on probability estimates. *Journal of Experimental Social

Psychology*, *42*(3), 380–387. https://doi.org/10.1016/j.jesp.2005.05.001

Röseler, L., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce judgment adjustment for both internally and externally generated anchors? *Journal of Experimental Social Psychology*, *87*, 103942. https://doi.org/10.1016/j.jesp.2019.103942

Röseler, L., Schütz, A., & Starker, U. (2019). Cognitive Ability does not and cannot Correlate with Susceptibility to Anchoring Effects. Advance online publication. https://doi.org/10.31234/osf.io/bnsx2

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2160588

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

Van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Wingen, T., Berkessel, J. B., & Englich, B. (2019). No Replication, No Trust? How Low

    Replicability Influences Trust in Psychology. *Social Psychological and Personality Science*,

    194855061987741. https://doi.org/10.1177/1948550619877412

Yeager, D. S., Bryan, C., & O'Brien, J. (2019). Replicator Degrees of Freedom Allow

    Publication of Misleading 'Failures to Replicate'. *SSRN Electronic Journal.* Advance online

    publication. https://doi.org/10.2139/ssrn.3408200

Article 3: Anchoring and Cognitive Ability

The following manuscript is the submitted version of a manuscript that is currently under review at the European Journal of Personality.

Röseler, L., Schütz, A., & Starker, U. (2019). Cognitive Ability does not and cannot Correlate with Susceptibility to Anchoring Effects. Advance online publication. https://doi.org/10.31234/osf.io/bnsx2

Cognitive Ability Does Not and Cannot Correlate with Susceptibility to Anchoring Effects

Lukas Röseler

Harz University of Applied Sciences, University of Bamberg

Astrid Schütz

University of Bamberg

Ulrike Starker

Harz University of Applied Sciences

This is a preprint (version 5). Since version 4, we added another study to the meta-analysis, updated the code for the meta-analysis, and fixed the cross-references. Changes since the version 3 include a meta-analysis of 15 effects across our study and four other studies, an overview of all reported reliabilities of anchoring tasks (in which we also included a study that was published this year), and a few minor corrections.

https://doi.org/10.31234/osf.io/bnsx2

Abstract

On a scale from 1 to 10, would you consider anchoring effects to be more or less important than a rating of 9? The seminal studies on anchoring effects are still surrounded by many mysteries regarding the question of how these effects come about. One of the mysteries is the influence of a person's cognitive ability. Some studies found that cognitive abilities moderated participants' susceptibility to anchoring, but others did not. We aimed to resolve these inconsistencies by making a distinction between informative and uninformative (e.g., random) anchors. In a preregistered online experiment, we tested the hypothesis that anchoring effects are weaker for people with higher cognitive abilities but that this relation only occurs if the anchors are perceived as random, and it vanishes if they are presented as informative. We found no support for the hypothesis. Results from a meta-analysis of 15 effects across our study and four other studies revealed no overall effect of cognitive ability on the susceptibility to anchoring, $g = 0.003$, 95% CI [-0.031, 0.037], $N_{total} = 1165$. Moreover, we observed that 10 items across three typical anchoring tasks had only very low internal consistency ($\alpha = .11$). Our analysis of additional published and unpublished data confirmed that a person's susceptibility to anchoring cannot be measured reliably. This explains why previous results on possible moderators of anchoring (e.g., studies on the Big Five or on fluid intelligence) have been highly inconsistent. We suggest that research on anchoring moderators needs to take a step back and develop a reliable measure for the susceptibility to anchoring.

Keywords: anchoring effect, cognitive ability, random anchors, reliability, meta-analysis

Cognitive Ability Does Not and Cannot Correlate with Susceptibility to Anchoring Effects

If people consider an evidently arbitrary numeric value before completing an estimation task, their estimate will usually be closer to that numeric value than when such a prime is not given. The phenomenon has been termed the *anchoring effect*.

Anchoring effects are well-known in decision making research. However, they are still not well understood, and numerous competing theories (e.g., Frederick & Mochon, 2012; Mussweiler & Strack, 1999) are aimed at explaining the phenomenon. The theories are nonexclusive, and none are sufficient for explaining the wide range of anchoring phenomena that have been observed over the last four decades. Moreover, replication issues have recently arisen for key findings from the Selective Accessibility Model (Harris et al., 2019) and the Numeric Priming account of anchoring (Klein et al., 2018).

Some studies have investigated the question of whether and how anchoring effects are associated with personality traits. There were some initial findings that showed moderating effects of conscientiousness, agreeableness, extraversion (e.g., Eroglu & Croxton, 2010), and openness to experience (e.g., McElroy & Dowd, 2007), but these could not be replicated (Furnham et al., 2012; for an overview, see also Cheek & Norem, 2019). Various studies have investigated cognitive ability as a possible moderator of susceptibility to anchoring. One study found that cognitive ability did indeed moderate susceptibility to anchoring (Bergman et al., 2010), whereas others found no evidence of such a moderating effect (Furnham et al., 2012; Stanovich & West, 2008; Teovanović, 2019). Briefly, findings on various moderators of anchoring effects have been inconsistent and contradictory, but the reason for these inconsistencies is still unknown.

Specifically for cognitive abilities, their effect on anchoring is inconclusive. Thus, we aimed to resolve the inconsistencies by embedding existing findings into Stanovich and West's (2008) framework, which conceptualizes individual differences in the operation of heuristics and thinking biases.

When does Cognitive Ability Influence Anchoring Effects?

Anchoring effects refer to the observation that numerical estimates are assimilated toward a previously presented value. We use the term in its plural form because of the large variety of paradigms that are used. Anchoring effects have been shown to be robust and to involve multiple nonexclusive mechanisms, such as insufficient adjustment and priming (e.g., Simmons et al., 2010; Chaxel, 2014) and scale distortion (Frederick & Mochon, 2012). Yet, individual differences in the susceptibility to anchoring effects are not clear, and with respect to the role of cognitive abilities, the findings are rather heterogeneous. Basing our work on Stanovich and West's (2008) framework, we show that additional factors need to be distinguished such as whether anchors are perceived as plausible estimates with respect to the issue at hand. In this framework, there are several conditions that must be met for cognitive ability to moderate whether there will be a heuristic response (i.e., strong anchoring effect) or a deliberate response (i.e., no anchoring effect). We argue that the crucial parameter is whether participants perceive that there is a need to override their initial heuristic response. We hypothesize that when people are confronted with anchors that are obviously useless with respect to the estimate, they know they need to ignore that initial anchor to provide an accurate estimate. In this case, cognitive ability may help them free their reasoning from the anchor. However, when people interpret the anchor as a useful hint, there is no need to override the initial response. In this case, cognitive ability should not make a difference. Past findings differ with respect to whether an effect of

cognitive ability on the susceptibility to anchoring effects was found (Bergman et al., 2010) or not (Stanovich & West, 2008). They also differ with respect to whether the anchor could be interpreted as informative (e.g., it was provided by the experimenter) or not (e.g., it was presented as a randomly generated number). In our study, we manipulated whether anchors were informative or random and expected cognitive ability to moderate the susceptibility to anchoring effects only when anchors were informative.

Method

Our hypothesis, method, and plan for data analysis were rigorously preregistered (https://osf.io/n8d6v/). We report how we determined our sample size, all data exclusions, all manipulations, and all measures (Simmons et al., 2012). We deviated from our preregistration in that we recruited more participants due to an initial lack of participants. This deviation did not affect the results. We manipulated whether the anchor that participants received was presented as a random number or as an informative suggestion. Our manipulation check consisted of two items that asked whether the number that was presented (i.e., the anchor) was determined randomly and whether participants tried not to let the number influence their estimate. The question about whether participants let the number influence their estimate was used for exploratory purposes to investigate whether participants readjusted their values consciously after they were presented with the anchor.

Participants

Data were collected online via social networks; participants were rewarded with feedback on their intelligence test. Besides the several types of media we planned to use (e-mail: via a student newsletter, university mailing lists, and social networks), we used additional sources to

achieve the sample size we aimed to reach. We posted our study on the web page of an online psychology journal and contacted trainee nurses via e-mail.

We had planned to recruit 174 participants up to January 31, 2018. Due to an initially low response rate, we extended the recruitment period and oversampled. The sample consisted of 225 complete cases (69.8% women) with a mean age of 26.54 years ($SD = 8.62$). This raised the power from .80 to .89. The results of the hypothesis test were the same no matter whether we used the full sample of 225 or only the first 174 participants. Participants were excluded from the analysis if at least one of the following preregistered criteria was met: (a) missing values, (b) overly consistent responses, (c) estimated celebrity ages in the age estimate task below 0 or above 100 on the age estimation task, (d) price estimates below 0€ or 0 cents and above 100€ or 1000 cents on the willingness to pay task, and (e) age above 100 years or below 18 years. The dropout rate was 724/949 clicks (76% drop out) of which most were excluded because they quit the study early.

Manipulations

Between-subject manipulations. The relevance of the anchor was determined by random computerized assignment to one of two groups: In the random anchor group, participants clicked on a box to generate a random anchor before each trial, whereas in the informative anchor group, participants were immediately presented with the anchor, which was described as a suggestion.

Within-subject manipulations. The anchor tasks consisted of three blocks of within-subject anchoring tasks with the following blocks and manipulations.

Quantities. The estimates of quantities featured six matrices with black or white boxes. They were created by a 2 (high vs. low anchor) × 3 (true value is 40, 50, or 60) complete within-

subject design. The assignment of a high or low anchor to the stimulus was completely randomized. High and low anchors were the true value plus and minus 30 (e.g., for the two matrices with 40/100 black squares, the anchors were 10 and 70). As this task involved perception rather than thinking, the influence of cognitive ability on susceptibility to anchoring could generally be lower than in other tasks as no actual reasoning was involved in the process of disregarding the anchor.

Prices. The estimates of prices (i.e., participants' *willingness to pay*; Bergman et al., 2010, p. 66) featured six products that were also created by a 2 (high vs. low anchor) × 3 (product: keyboard, apple, pen) design. The two keyboards, apples, and pens were chosen to visually resemble each other. Using a pilot test, we had made sure that they were perceived as equally valuable. The high and low anchors were (in the same order as the products) 20€ versus 10€, 80ct versus 20ct, and 90ct versus 50ct, respectively.

Age. The age estimation task was chosen as a general knowledge test and featured eight people who were also chosen according to a 2 (high vs. low) × 4 (age of the person) design. The respective people were pairs of celebrities who were actually the same age. To determine the anchors, their age was rounded to the nearest 10, and then 10 years were added (high anchor) or subtracted (low anchor). The people we chose were: Joschka Fischer (German politician), Al Gore (U.S. politician), Sandra Bullock (actress), Russel Crowe (actor), Jessica Alba (actress), Fernando Alonso (Formula 1 driver), Lionel Messi (soccer player), and Sebastian Vettel (Formula 1 driver). Using a pilot test, we had made sure that the age estimates for the matched celebrities did not differ and that people were usually familiar with them but did not know their actual age. On the list, two consecutive people had the same age at the time of the experiment. They were presented with their name and profession.

Measures and Materials

Cognitive ability was measured with the Hagen Matrices Test (HMT; Jäger & Heydasch, 2016). Self-control was measured with a German version of the Self-Control Scale (Sproesser et al., 2011; Tangney et al., 2004). Self-esteem was measured with the German version of Rosenberg's Self-Esteem Scale (Collani & Herzberg, 2003). Susceptibility to anchoring effects was measured with participants' reactions to the tasks described above. We chose to include a new task (the quantity estimates task) because it has several advantages over commonly used anchoring tasks: (a) the existence of maximum, minimum, and true values (which are unknown to participants); (b) the anchors can be used symmetrically on two consecutive stimuli with the same true value (e.g., for a true value of 50, the anchors are 50±30); (c) The repeated-measures manipulation helps to test for the reliability of susceptibility scores and provides high power; (d) the random and difficult-to-remember stimuli allow the tasks to be repeated. For example, if we asked about the boiling point on Mt. Everest (e.g., Epley & Gilovich, 2001), people who did not know the answer might look it up after the experiment and know that it is 71°C (160 °F). This approach addresses a big problem of tasks such as the cognitive reflection task (Frederick, 2005), which is that the measure's construct validity deteriorates when participants know the answer or look it up (Toplak et al., 2013). The repeated-measures design allowed us to examine whether susceptibility to anchoring effects is robust across multiple anchoring tasks, and we could thus estimate its reliability.

Procedure

First, participants answered demographic questions (i.e., sex, age) and provided a personal code for potential subsequent experiments (i.e., to assess the retest reliability of the susceptibility-to-anchors score). Then they completed a brief version of the Self-Control Scale

(German translation) and the Rosenberg Self-Esteem Scale. Next, participants were randomly assigned to one of two experimental conditions. In each condition, the participants performed three anchoring tasks: (a) a quantity estimation task, (b) an age estimation task, and (c) a willingness to pay task. Half of the participants were instructed that they would be presented with a random number. The numbers 10, 20, …, 90 were presented for 50 ms in a random order. Participants pressed a "generate random number" button and were then given the anchor. The other half of the group was given a number that the experimenter proposed to them.

For the quantity estimation task, the participants were instructed that they would see six 10x10 matrices with black and white squares for 1 s each. They were asked to estimate the number of black squares after each presentation. The task was followed by the age estimation task and the purchase decision.

Then, participants were asked whether they thought that they were influenced by the number they had been given and whether they thought the number was determined randomly (manipulation check). For exploratory purposes, we asked whether they had tried to suppress thoughts of the number, whether they had their estimate in their minds before they were presented with the number, and whether they had heard about the anchoring effect.

Then, participants conducted the HMT (test of fluid intelligence). The test was presented at the end due to technical restrictions. We believe that the measurement time should not be relevant as intelligence is considered a trait, and performance should be independent of the tasks conducted beforehand.

Results

Reliability of Measures

The internal consistency of the measure of cognitive ability (HMT) was α = .75 (20 items, N = 225). Internal consistency was excellent for the Self-Control Scale (α = .84, 13 items) and the Rosenberg Self-Esteem Scale (α = .91, 10 items). The general tendency to be susceptible to anchoring was measured on the basis of the correlations between the three tasks. The results showed very low correlations between the three task scores (α = .13, α = .11 for all 10 items; see Table 1). The probability that 3 of 45 correlations would be significant given H0 is p = .492.

Table 1

Correlations between Susceptibility to Anchoring Scores

	Score	1	2	3	4	5	6	7	8	9	10
1	Quantity 50	1	.185**	.052	-.088	.054	.053	.136*	-.044	-.101	.061
2	Quantity 40	.185**	1	.087	.058	.091	.051	-.038	.029	.022	.024
3	Quantity 60	.052	.087	1	.111	.076	-.028	.028	-.048	-.026	.025
4	Age 69	-.088	.058	.111	1	-.078	.008	.093	-.002	-.007	.010
5	Age 53	.054	.091	.076	-.078	1	.054	-.114	-.110	.049	-.001
6	Age 36	.053	.051	-.028	.008	.054	1	.059	.020	.125	-.042
7	Age 30	.136*	-.038	.028	.093	-.114	.059	1	.057	.158*	-.002
8	WTP pen	-.044	.029	-.048	-.002	-.110	.020	.057	1	.101	-.007
9	WTP apple	-.101	.022	-.026	-.007	.049	.125	.158*	.101	1	-.003
10	WTP keyboard	.061	.024	.025	.010	-.001	-.042	-.002	-.007	-.003	1

Note. N = 225.

*p < .05 (two-tailed). **p < .01 (two-tailed).

Manipulation Checks and Data Quality Assessment

Randomness. Participants who had "generated" their random anchor by clicking a "stop" button next to a fluctuating number (random anchor condition) perceived the anchor as more random (M = 2.93, SD = 1.22, N = 104) than participants who were presented with the anchor by

the experimenter (informative anchor condition; $M = 2.40$, $SD = 1.16$, $N = 121$). The

manipulation check was significant with a medium-sized effect, $t(223) = 3.38$, $p < .001$, $d = 0.45$.

Exploratory items (not preregistered). Participants in the random anchor condition felt

that the anchor exerted a smaller influence on them than participants in the informative anchor

condition ($M_{random\ anchor} = 2.88$, $SD = 1.07$, $N = 104$; $M_{informative\ anchor} = 3.02$, $SD = 1.08$, $N = 121$).

However, this difference was not statistically significant, $t(223) = -0.98$, $p = .330$, $d = 0.13$.

There was no statistically significant difference between the two experimental conditions

concerning the extent to which participants tried not to let themselves be influenced by the

anchor ($p = .522$). Moreover, styles of estimating differed between groups. Participants in the

informative anchor condition more often reported that they had thought of a number before they

saw the anchor ($M_{random\ anchor} = 2.24$, $SD = 1.22$, $N = 104$; $M_{informative\ anchor} = 2.88$, $SD = 1.17$, $N = 121$). The effect was medium in size, $t(223) = -3.99$, $p < .001$, $d = 0.54$.

Strength of anchoring effects in general (exploratory preliminary analyses). All but

two out of 10 items showed significant anchoring effects, namely, the price estimations for the

pen ($p = .243$; one-sided test) and the keyboard ($p = .090$; one-sided test). The mean effect size

for the anchoring effects was $dz = .26$ ($dz = .31$ when the nonsignificant items were excluded).

For an overview of the anchoring effects, see Figure 1.

Figure 1

Estimates per task and anchor reveal anchoring effects for all but two items
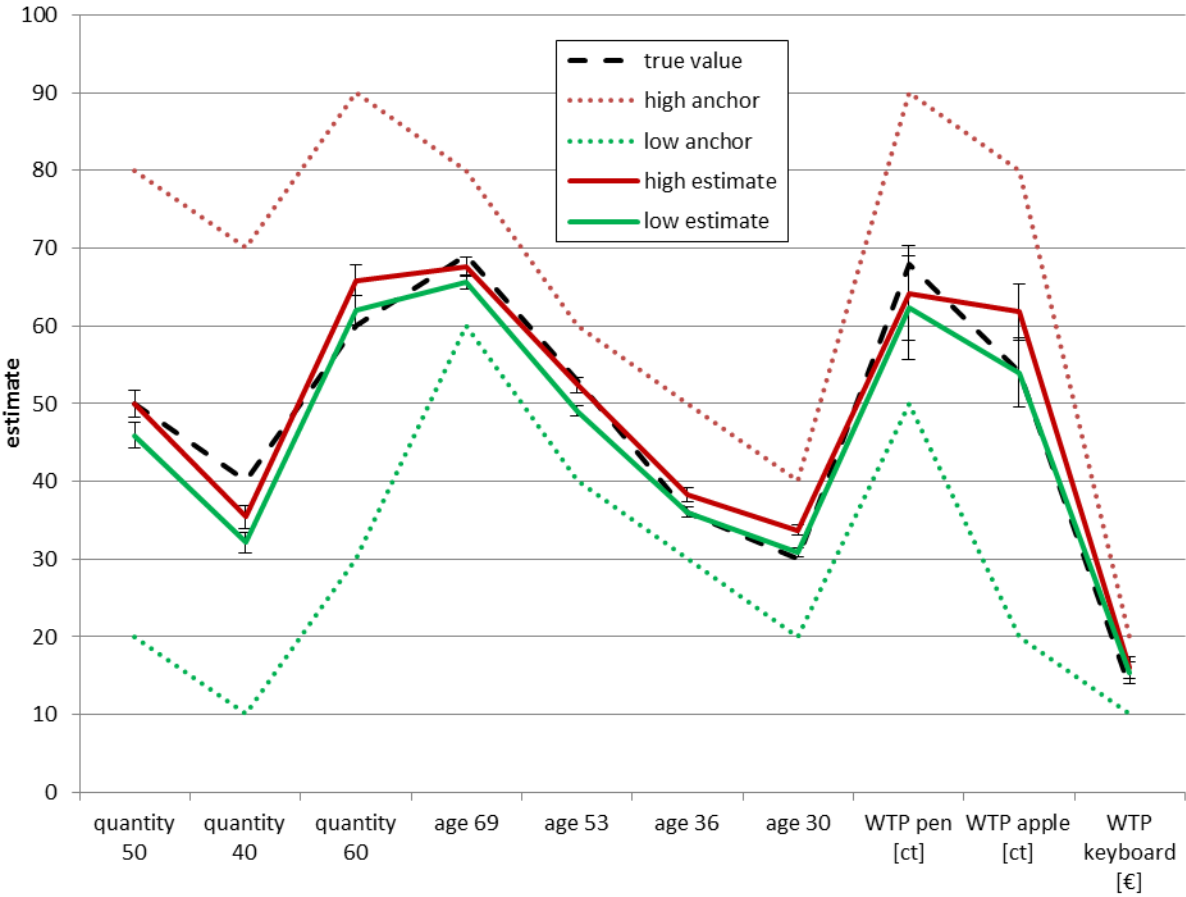
Note. Error bars indicate 95% confidence intervals.

Calculation of Scores

Cognitive ability was measured as the number of tasks solved on the HMT (i.e., 0 to 20). Susceptibility to anchoring effects was computed differently for the three tasks. In the quantity estimation tasks, we computed the z-value for the difference between the mean of the high anchor estimates and the mean of the low anchor estimates. In the willingness to pay task (price estimation), we computed the standardized mean values (i.e., z-values) for the difference between the high anchor price estimate and the low anchor price estimates. Note that the computation was different for the products because the ranges of the anchors were not constant (unlike in the other tasks). In the age estimation task, we used the standardized mean values (i.e., z-values) for the difference between the high anchor estimates and the low anchor estimates.

To test our hypothesis, we used a MANOVA with a complete design. The dependent variables were the susceptibility to anchoring effects scores for the three tasks (quantity estimation task, age estimation task, and willingness to pay task). The independent variables were cognitive ability and the relevance of the anchor. We expected an interaction between cognitive ability and the relevance of the anchor with an effect size of $f = 0.214$ or $\eta^2 = .043$. We did not expect any of the other effects in the model to be significant.

Hypothesis Test

Contrary to our preregistered hypothesis, the interaction between cognitive ability and the relevance of the anchor had no effect on susceptibility to anchoring effects, $F(1, 221) = 1.37$, $p = .243$, $\eta_p^2 = .006$. Instead, there were interactions between the task and cognitive ability, $F(2, 442) = 4.84$, $p = .008$, $\eta_p^2 = .021$, as well as between the task and the relevance of the anchor, $F(2, 442) = 3.33$, $p = .037$, $\eta_p^2 = .015$. Furthermore, the main effect of relevance was significant, $F(1,$

221) = 4.66, $p = .032$, $\eta_p^2 = .021$. All other effects were not significant. All analyses reported hereafter were exploratory and were not preregistered.

As recommended by Cheek and Norem (2018), we conducted the same analysis as above using the absolute differences between anchors and estimates as the strength of the anchoring effect. This did not change the results. None of the effects described in the model above were significant except for the anchor item (i.e., absolute differences were larger for some items than for others). The internal consistency of all 20 items (two times the 10 items described above because each was presented with a low and high anchor) was slightly larger but still far from acceptable, $\alpha = .385$. There was no correlation between cognitive ability and the mean of all absolute differences, $r(223) = -.020$, $p = .761$.

Post hoc analyses (i.e., not preregistered analyses) revealed that cognitive ability was more strongly correlated with susceptibility to anchoring effects for the score on the quantity estimation task, $r(223) = -.24$, $p < .001$, than for the scores on the age estimation task, $r(223) = -.05$, $p = .459$, or the willingness to pay task, $r(223) = .06$, $p = .399$. We conducted two-tailed z-tests between the three correlations indicating that only the differences between the quantity task and the other two tasks were significant ($z_{quantity\ vs.\ age} = -2.06$, $p = .039$, $z_{quantity\ vs.\ wtp} = -3.18$, $p = .002$, $z_{age\ vs.\ wtp} = -1.12$, $p = .263$). Note that this analysis was not part of our hypothesis and that when we applied a Bonferroni correction to protect against making a Type 1 error, the first difference became nonsignificant.

As our score for a low susceptibility to anchoring effects was sensitive to accurate estimates of the stimuli, we checked whether cognitive ability was correlated with accuracy to eventually control for it. For the accuracy score, we computed the sum of absolute deviations from the actual value. High cognitive ability was correlated with a low absolute deviation from

the true value (i.e., the true number of black boxes), $r(225) = -.17$, $p = .009$. As the relationship between a low susceptibility to anchoring effects and cognitive ability could be triggered by the accuracy effect, we computed the partial correlation between cognitive ability and susceptibility to anchoring effects while controlling for accuracy and found that the correlation barely changed, $r(222) = -.22$, $p < .001$. The main effect of the relevance of the anchor hinted that the source was important after all. Informative anchors are stronger than overall anchors, $t(223) = 2.75$, $p = .006$, $d = 0.37$.

Self-control and self-esteem. To explore whether personality traits such as self-control and self-esteem have an impact on people's susceptibility to anchoring effects, we computed correlations for participants' self-control and self-esteem with their susceptibility to anchoring effects. Neither variable was significantly correlated with susceptibility: self-control, $r(223) = -.02$, $p = .753$; self-esteem, $r(223) = .03$, $p = .627$.

Knowledge of the anchoring effect. People who were familiar with the anchoring effect were less affected by the anchors ($M = -.22$, $SD = .61$, $N = 65$) than people who were not familiar with it ($M = .09$, $SD = .59$, $N = 165$), $t(224) = 2.99$, $p = .003$. Note that the two groups did not differ in their cognitive abilities ($M_{yes} = 8.82$, $SD = 3.84$, $N = 65$; $M_{no} = 8.08$, $SD = 3.41$, $N = 160$), $t(223) = 1.41$, $p = .160$, $d = 0.20$. More specifically, absolute differences between the high anchor and low anchor estimates for each of the pairs of stimuli were all lower for people who knew of the anchoring effect. However, only one of the pairwise comparisons was significant ($p = .049$).

Discussion

In this research, we applied Stanovich and West's (2008) framework, which conceptualizes individual differences in the operation of heuristics and thinking biases, to people's susceptibility to anchoring effects with the aim of explaining why different studies have found different results concerning the relation between cognitive ability and susceptibility to anchoring effects. We aimed to test the relevance of anchors (presenting them as either random or informative) as a crucial variable that might determine whether cognitive ability does or does not weaken the influence of anchors.

One of our primary goals was to explain why Bergman et al. (2010) found a moderating effect of cognitive ability in their experiment, whereas in the experiments conducted by Stanovich and West (2008) and Furnham et al. (2012), no moderating effects were found. However, we are still not able to provide an explanation on the basis of our results. Instead, our results add to the contradictions and inconsistencies: We did not find a moderating effect on the general knowledge (age estimation) task, thus contradicting Stanovich and West (2008) and Furnham et al.'s (2012) results. We also did not find an effect on the willingness to pay task, thus contradicting Bergman et al.'s (2010) results. The most likely explanation for our findings is that susceptibility to anchoring effects cannot be measured reliably. Although the phenomenon of anchoring is said to be robust—and sometimes even termed *reliable*—there is apparently yet no way to measure susceptibility to anchoring *reliably* on a participant level.

Meta-Analysis of the Effect of Cognitive Ability on Susceptibility to Anchoring

Table 2 provides an overview of all experiments involving cognitive ability and anchoring effects. Apart from the presentation of the stimulus, there is currently no single factor that can explain the presence or absence of a correlation between cognitive ability and the

susceptibility to anchoring effects. Given the low reliability, the two positive effects are puzzling.

Note that Bergman et al. (2010) reported a p-value of $p = .043$ (p. 67) and Teovanović (2019) who only found an effect after splitting the sample into high and low scores based on presumptions about subjects' reflectivity, their score in the cognitive reflection task ($p = .003$, p. 16). To determine whether the overall effect is homogeneous and significantly larger than zero, we conducted a meta-analysis. Due to insufficient reports, we had to assume $F = 1$ for cases in which $F < 1$ was reported (Furnham et al., 2012). A multilevel random effects meta-analysis revealed no significant effect of all 15 effects across the 5 studies, $t(15) = 0.22$, $p = .827$, $g = 0.004$, 95% CI = [-0.030, 0.037], $N_{total} = [1435, 1448]$. Effects were homogeneous, $Q(15) = 3.35$, $p = .999$ (data and code are available at https://osf.io/9wnd4/). See Figure 2 for a forest plot of the results.

Table 2

Designs and Results for all Studies Involving the Susceptibility to Anchoring Effects and Cognitive Ability

Variable	Bergman et al., 2010	Stanovich & West, 2008	Furnham, Boo, & McClelland, 2012	Teovanovic, 2019	Quantity Estimation	Age estimation	Willingness to Pay Estimation
Task	Willingness to pay with "real decisions"	General knowledge	General knowledge	General knowledge	Quantity estimation	General knowledge	Willingness to pay with hypothetical decisions
Source of the anchor	Social Security Number	Determined by experimenter	Determined by experimenter	Determined by experimenter	Random or determined by experimenter	Random or determined by experimenter	Random or determined by experimenter
Cognitive ability measure	Predicting Job Performance [mainly fluid intelligence]	SAT score [crystallized intelligence]	Baddeley Reasoning Test (1968) [fluid intelligence]	Raven's matrices (1979) [fluid intelligence]	HMT [fluid intelligence]	HMT [fluid intelligence]	HMT [fluid intelligence]
Calculation of susceptibility to anchoring scores	None	None	Logarithmic transformation	Difference between no-anchor estimate and anchor estimate	Standardized by item and task	Standardized by item and task	Standardized by item and task
Method	Correlation	Interaction in a linear model	Interaction in a linear model	Correlation	Correlation	Correlation	Correlation
Participants' country of origin	Sweden	USA	USA	Serbia	Germany	Germany	Germany
High anchors	[00, 99]	65, 1000	68, 77, 113, 187	Computed individually	80, 70, 90	80, 60, 50, 40	90, 80, 20
Low anchors	[00, 99]	12, 85	22, 17, 49, 24		20, 10, 30	60, 40, 30, 20	50, 20, 10
Mean anchoring effect size [standardized r]	.369	.581, .721	.457	[.206, .242]	.207	.289	.083
Kind of stimulus	Presented and disguised OR visible during estimation	Remembered from LTM	Remembered from LTM	Remembered from LTM	Presented, then disguised	Remembered from LTM	Visible during estimation
Sample size	116	434	172	236	225	225	225
Design of factor "anchor"	Between subjects	Between subjects	Between subjects	Within subjects	Within subjects	Within subjects	Within subjects

Figure 2

Forest plot and prediction interval for all effects of cognitive ability on strength of anchoring effects

Study	Anchoring and Cognitive Ability	SMD	95%-CI
Bergman et al., 2010		0.48	[0.11; 0.86]
Stanovich & West, 2008, African Countries		0.16	[-1.84; 2.15]
Stanovich & West, 2008, Redwoods		0.00	[0.00; 0.00]
Furnham et al., 2012, General Intelligence A		0.09	[-0.32; 0.50]
Furnham et al., 2012, General Intelligence B		0.00	[0.00; 0.01]
Furnham et al., 2012, General Intelligence C		0.00	[0.00; 0.01]
Furnham et al., 2012, General Intelligence D		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning A		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning B		0.00	[0.00; 0.01]
Furnham et al., 2012, Verbal Reasoning C		0.09	[-0.32; 0.50]
Furnham et al., 2012, Verbal Reasoning D		0.00	[0.00; 0.01]
Teovanovic, 2019, Raven's Matrices		0.06	[-0.03; 0.15]
Teovanovic, 2015, Raven's Matrices		0.06	[-0.18; 0.30]
Röseler et al., 2019, Quantity Estimates		0.49	[0.24; 0.75]
Röseler et al., 2019, Age Estimates		0.10	[-0.16; 0.36]
Röseler et al., 2019, Price Estimates		0.02	[-0.22; 0.26]

Hedge's g

Low Reliability of the Susceptibility to Anchoring Effects in Other Data Sets

To even more thoroughly test our claim about the lack of reliability in the anchoring effects measures as an explanation of no correlations and inconsistent results, we computed the correlations between different anchoring tasks in other data sets. We analyzed some previous data from our own lab that involved anchoring effects in general knowledge questions with a sample of 650 participants. We were able to confirm the low reliability ($\alpha = .05$, 16 items).

We further analyzed the data provided by Cheek and Norem (2019; https://osf.io/a6ywr/), which had been used to test for interactions between the Big Five and the susceptibility to anchoring effects. We computed the absolute differences between the anchors, standardized

them, excluded outliers as they did (±5 SD), and again found an unacceptable internal consistency (six items) of α = .14 and a maximum correlation between the two susceptibility to anchoring effects scores of r = .083.

Another data set was provided by Bergman et al. (2010), who found an interaction between susceptibility to anchoring effects and cognitive ability. The anchor in this case was the last two digits of participants' social security numbers and was used for six product price estimates. As in the other data sets, we computed the absolute differences between the anchor and the price estimates and found an internal consistency (six items) of α = .69. This was considerably higher than in the other data sets. Given that the anchor was the same for all six items, and all six tasks were actual price estimates (nonhypothetical decisions), the shared variance between the items may, however, also reflect a generally higher willingness to pay. It is possible that the measure used by Bergman et al. (2010) is not as a measure of susceptibility to anchoring effects but is at least partly a measure of willingness to pay. This would make their significant correlations between cognitive ability and the product price estimates difficult to interpret.

Similarly to Bergman et al. (2010), Frech et al. (2020) also used price estimates. Additionally, they only used high anchors and measured stepwise adjustment from those anchors, thus confounding susceptibility to anchoring with overall willingness to pay. Internal consistency was α = .513 (3 items) and thus larger, too (see https://osf.io/2fkpm/ for the analysis code and https://osf.io/dq48j/ for the original dataset).

Finally, we found a publication in which the average of the correlations between the anchoring items was reported. Similar to our approach, Teovanović (2019) used a within-subjects anchor (no anchor vs. anchor) manipulation for the general knowledge questions and

found an average correlation of $r = .11$ or $r = .27$ depending on how it was calculated and $\alpha = .71$ (8 items, p. 15). It is unclear why this correlation is much higher than correlations between anchoring items in other studies.

Given the wide range of tasks used in our own study and the occurrence of the problem in other studies, we believe that the unreliable measurement of the susceptibility to anchoring effects poses serious problems for research on anchoring effect moderators. An overview of reported and computed reliabilities is provided in Table 3. Cronbach's alphas are significantly different between studies, $\chi(5) = 126.17$, $p < .001$.

Table 3

Reliabilities available from anchoring research.

Study	Coefficient Alpha	Alpha LCB	Alpha UCB	Average r	Number of Items	N	Note
Teovanovic, 2019	0.75	0.70	0.79	0.11	24	236	
Bergman et al., 2010	0.69	0.59	0.77	0.27	6	116	Confounded with WTP
Frech et al., 2020, Study 2	0.51	0.38	0.62	0.26	3	189	Confounded with WTP
Cheek & Norem, 2019	0.14	0.05	0.22	0.03	6	945	
This study	0.11	-0.07	0.27	0.01	10	225	
Röseler et al., 2020	0.05	-0.06	0.15	0.00	16	650	

Note. LCB = Lower Confidence Bound, UCB = Upper Confidence Bound for $\alpha = .05$, WTP = Willingness to Pay.

Use of Difference Scores as a Potential Aspect of the Problem of Low Reliability

Part of the problems encountered with unreliable measurements may be due to the use of difference scores. That is, the reliability of the difference score (e.g., the estimate used for the high anchor minus the estimate used for the low anchor) depends on the reliabilities of the scores from which the difference is computed and the correlation between the scores from which the difference is computed (e.g., Peter et al., 1993). The difference between two unreliable scores is not reliable, and the difference between two strongly correlated scores (e.g., $r = 1$) has no

variance and is not reliable either. Parsons et al. (2018) noted that the reporting of reliability in cognitive bias tasks should be based on a standard to allow for an assessment of how widespread the problem of low reliabilities is. The problem is already known for the dot-probe task, where differences between reaction times are computed to measure attentional biases. Reports of reliability ranged from $r = -.23$ to $r = .70$ (Parsons et al., 2018, p. 15). Due to a lack of reporting standards for tasks such as anchoring, it is unknown how widespread the problem of low reliability is. Anchoring effects are clearly no exception to both the problem of missing reporting standards and the problem of low reliability.

It is possible that there are some items for which intercorrelations may be high for different reasons. For these cases, we advise researchers to critically examine whether the items may measure constructs beyond the susceptibility to anchoring effects. For example, using willingness to pay and high anchors may only yield strong anchoring effects for participants who have a high general willingness to pay. However, this may hold not because anchoring effects are stronger for them but because they have a higher income and give higher "estimates" on average.

Applicability of the Framework for Conceptualizing Individual Differences on Heuristics and Biases Tasks

Because the data were not reliable, we were unable to test the hypothesis we derived from Stanovich and West's framework. Although this is a serious limitation, we uncovered a very important problem instead. We would even go so far as to say that it is not yet possible to test the hypothesis we derived from the framework. To test the model—and in fact, to test any interaction model on anchoring effects—a reliable assessment of participants' susceptibility to anchoring effects is necessary.

Future Directions for Personality Trait Research in the Context of Anchoring Effects

To date, findings on moderators of anchoring effects (e.g., correlations between personality or intelligence with people's susceptibility to anchoring effects) are still inconclusive. For example, positive, negative, and null correlations have all been found between each of the Big Five and susceptibility to anchoring effects (Cheek & Norem, 2019). Our results show that anchoring effects cannot be moderated by any trait because the measure of susceptibility is not reliable. Given the unreliable assessment of susceptibility to anchoring effects, our null effects are not surprising and neither are the contradictory findings in the literature. With one exception, past studies did not report the reliability of the anchoring measures, although most designs allowed for a test of reliability, as we demonstrated using the data sets provided by Cheek and Norem (2019) and Bergman et al. (2010). As long as the reliability problem remains, any null findings for anchoring effect moderators do not indicate the absence of an actual effect, and any positive findings require close inspection. Other biases (e.g., the dot-probe task) face similar problems (Parsons et al., 2018), which could partly be resolved through theoretical work (Schäfer et al., 2016). Increasing the reliability of anchoring effects might be possible through the development of more specific theories.

For example, formalization of anchoring effects with the aid of cognitive models such as SARA (Selective Activation, Reconstruction, and Anchoring; Pohl et al., 2003) or TOTE-units (test-operate-test-exit; Epley & Gilovich, 2001) could enhance the reliability of anchoring measures. A descriptive approach was described by Turner and Schley (2016). Given that dynamic models have been used to solve reliability issues for other effects, this research strategy currently seems the most promising in the search for a reliable measure of the susceptibility to anchoring effects. Finally, we have to consider the possibility that observed differences in the

susceptibility to anchoring effects might not be due to actual individual differences but that there

might actually be no trait that can be identified as *susceptibility to anchoring.*

References

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

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

Cheek, N. N., & Norem, J. K. (2018). On Moderator Detection in Anchoring Research:

Implications of Ignoring Estimate Direction. *Collabra: Psychology*, *4*(1), 12.

https://doi.org/10.1525/collabra.125

Cheek, N. N., & Norem, J. K. (2019). Are Big Five Traits and Facets Associated With

Anchoring Susceptibility? *Social Psychological and Personality Science*, *92*,

194855061983700. https://doi.org/10.1177/1948550619837001

Collani, G. von, & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala

zum Selbstwertgefühl von Rosenberg. *Zeitschrift Für Differentielle Und Diagnostische

Psychologie*, *24*(1), 3–7. https://doi.org/10.1024//0170-1789.24.1.3

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

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts:

The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

https://doi.org/10.1016/j.ijforecast.2009.02.005

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. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology. General*, *141*(1), 124–133. https://doi.org/10.1037/a0024006

Furnham, A., Boo, H. C., & McClelland, A. (2012). Individual Differences and the Susceptibility to the Influence of Anchoring Cues. *Journal of Individual Differences*, *33*(2), 89–93. https://doi.org/10.1027/1614-0001/a000076

Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., Urlichich, D., & Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology. General.* Advance online publication. https://doi.org/10.1037/xge0000644

Jäger, M., & Heydasch, T. (2016). Hagener Matrizen-Test (HMT) (PSYNDEX Tests Review). In *Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Ed.), PSYNDEX. Trier: ZPID.*

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., Babalola, M. T., Bahník, Š., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Bocian, K., Brandt, M., Busching, R., Cai, H., Cambier, F., . . . Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. Advance online publication. https://doi.org/10.31234/osf.io/9654g

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making*, *2*(1), 48–53.

Mussweiler, T., & Strack, F. (1999). 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

Parsons, S., Kruijt, A.-W., & Fox, E. (2018). Psychological Science needs a standard practice of reporting the reliability of cognitive behavioural measurements. Advance online publication. https://doi.org/10.17605/OSF.IO/6KA9Z

Peter, J. P., Churchill, J. G. A., & Brown, T. J. (1993). Caution in the Use of Difference Scores in Consumer Research. *Journal of Consumer Research*, *19*(4), 655. https://doi.org/10.1086/209329

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). Sara: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory (Hove, England)*, *11*(4-5), 337–356. https://doi.org/10.1080/09658210244000487

Schäfer, J., Bernstein, A., Zvielli, A., Höfler, M., Wittchen, H.-U., & Schönfeld, S. (2016). Attentional Bias Temporal Dynamics Predict Posttraumatic Stress Symptoms: A Prospective-Longitudinal Study among Soldiers. *Depression and Anxiety*, *33*(7), 630–639. https://doi.org/10.1002/da.22526

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2160588

Sproesser, G., Strohbach, S., Schupp, H., & Renner, B. (2011). Candy or apple? How self-control resources and motives impact dietary healthiness in women. *Appetite*, *56*(3), 784–787. https://doi.org/10.1016/j.appet.2011.01.028

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Teovanović, P. (2019). Individual Differences in Anchoring Effect: Evidence for the Role of Insufficient Adjustment. *Europe's Journal of Psychology*, *15*(1), 8–24. https://doi.org/10.5964/ejop.v15i1.1691

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology*, *90*, 1–47. https://doi.org/10.1016/j.cogpsych.2016.07.003

Article 4: Reliable Measurement of Susceptibility to Anchoring

The following manuscript is the submitted version of a manuscript that is currently under review at the Journal of Personality.

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

The Stepwise Anchoring Paradigm: Measuring Reliable Components of Anchoring and Adjustment as the Next Step in Moderator Research

Lukas Röseler[1,2], Astrid Schütz[2], Ivonne K. Dolling[2], Kathrin Friedinger[2], Yolanda Hösch[2], Julia C. Hügel[2], Elisa L. E. Igna[2], Alexandra H. Karg[2], Sarah Kühling[2], Dominik Schäffer[2], Anke Zhou[2], and Jessica J. Röseler[2]

[1]Harz University of Applied Sciences

[2]University of Bamberg

Author Note

Abstract

Although anchoring research has explored personality moderators for at least 13 years, it has not produced any robust findings. The reason for this surprising fact is rooted in the problem that most anchoring tasks have unacceptable reliability levels. In a preregistered experiment, we introduce a paradigm that we call the *stepwise anchoring paradigm*. It is based on the incumbent insufficient adjustment model of anchoring and previous research by Frech et al. (2020). Using the stepwise anchoring paradigm and estimates of animals' weights, we reliably measured parameters of anchoring effects and investigated whether these parameters were correlated with theoretically relevant personality traits (i.e., cognitive reflection and impulsiveness). The reliability of the number of adjustment steps was acceptable ($\alpha = .749$, 12 items, $N = 123$) and much higher than the reliability of the total adjustment ($\alpha = .399$), which has previously been used as an indicator of susceptibility to anchoring. However, we found no correlations between anchoring parameters and impulsivity or cognitive reflection. Thus, a method for reliably measuring the susceptibility to anchoring is now available, but personality traits that may be related to this variable have yet to be identified.

Keywords: anchoring and adjustment, anchoring effect, reliability, cognitive reflection, impulsiveness

The Stepwise Anchoring Paradigm: Measuring Reliable Components of Anchoring and

Adjustment as the Next Step in Moderator Research

When people make numerical estimates, their estimates have been shown to be biased toward random numbers that they have previously considered. This phenomenon is called *anchoring* and has been investigated rigorously since it was made famous by Tversky and Kahneman (1974). Many models have been proposed to explain why and when anchoring occurs (e.g., Frederick & Mochon, 2012; Mussweiler & Strack, 1999a; Wegener et al., 2001), and in recent years, substantial research has been devoted to answering the question of which personality traits moderate the susceptibility to anchoring (e.g., Cheek & Norem, 2018; Furnham et al., 2012; Teovanović, 2019; Welsh et al., 2011).

However, after a decade of research on individual differences in anchoring, results have been extremely inconsistent. For example, Cheek and Norem (2019) found evidence of positive, negative, and null correlations between openness to experience and anchoring, and the same has been true for the relation between agreeableness and anchoring. The most plausible explanation for these blatant contradictions is the lack of a method to reliably assess people's susceptibility to anchoring (Röseler et al., 2019). In fact, how strongly an anchor influences people's judgments of one item is not or is only weakly correlated with how strongly an anchor influences their judgments of another item (for an overview of reliabilities of anchoring tasks, see Röseler et al., 2019, p. 21).

All Models Are Wrong, but Some Are Useful

Anchoring research that has been conducted without a solid theoretical foundation has not yielded many insights. For example, studies on incidental anchoring (i.e., presenting anchors below the attention threshold; Critcher & Gilovich, 2008) and subliminal anchoring (i.e.,

presenting anchors below the perception threshold; Mussweiler & Englich, 2005) could not be replicated (e.g., Klein et al., 2018; Röseler, Schütz, Blank et al., 2020, respectively). Thus, the most promising way to both develop an anchoring task that has high reliability and find moderators on this basis seems to refer to theories and models of anchoring.

However, this is easier said than done. There are many theories, and in line with Box (1976), each of them has been faced with findings that the theory has not been able to explain. Moreover, there have been contradictions, even among predictions that were derived from the same theory. Replication failures and a lack of empirical content in the domain of personality moderators have made strong theorizing even more difficult. Most researchers have not tried to falsify any of the models but have simply attempted to make their own model more plausible than the other models, for example, by stating that "… the explanatory sufficiency of each theory reduces the value of alternate theories…" (Mochon & Frederick, 2013, p. 77; see also Yarkoni, 2020, May 6 for a general discussion about falsification in psychological science). In the following, we summarize the most popular anchoring theories and argue for why the insufficient adjustment model (e.g., Epley & Gilovich, 2001) is the most useful theory for personality moderator research.

Selective Accessibility Model

According to the selective accessibility model (e.g., Mussweiler & Strack, 1999a, 1999b), people test whether the anchor is a good value (hypothesis-consistent testing) and activate knowledge that is consistent with the anchor (semantic priming). Potential moderators in this line of reasoning would be those that are relevant to semantic priming, such as need for cognition (Petty et al., 2008) or associative thinking (Wood et al., 2011). Although this model has been the most prominent one for some years, recent findings have raised doubts about its usefulness. For

example, in the case of ego depletion, different researchers derived from the selective accessibility model the opposing ideas that ego depletion leads to stronger (Banker et al., 2017) and weaker (Francis et al., 2018) anchoring effects, and both were wrong (Röseler, Schütz, Baumeister, & Starker, 2020). Even stronger doubts arose when the "most compelling evidence" for the selective accessibility model could not be replicated: No evidence for a faster recognition of words that were primed by an anchor (e.g., "summer" for a hot temperature anchor) could be found in a series of five preregistered experiments (Harris et al., 2019).

Scale Distortion Theory

The scale distortion theory is one of the more recent accounts of anchoring effects. In its original version, anchoring was construed as only a language effect (in contrast to a representation effect; Frederick & Mochon, 2012). Accordingly, in this line of reasoning, a high anchor does not increase the accessibility of high numbers as proposed by the Selective Accessibility Model, but it makes numbers that would otherwise seem plausible "appear smaller." For example, "1,000 pounds [seems more plausible as an estimate for] a larger animal in the presence of this small anchor than without it" (Frederick & Mochon, 2012, p. 2). A recent refinement of the model (Mochon & Frederick, 2013) yielded two conditions that must be met for anchoring to occur: (a) the anchor and target must be assessed on the same scale and (b) the anchor and target must be related, or respondents must be forced to compare them. Mochon and Frederick (2013) argued that random anchors actually carry conversational norms (Grice, 1982). However, this account of anchoring is not an exclusive one—that is, it does not contradict other models.

When asking how assimilation toward the anchor actually works, the model does not offer a detailed account. Thus, to derive specific predictions, one has to rely on other models.

Furthermore, hypotheses about which personality traits (besides the knowledge of conversational norms) may be relevant moderators cannot be derived from scale distortion theory.

Insufficient Adjustment Model

A more specific model is the insufficient adjustment model. This model was inspired by Tversky and Kahneman's (1974) brief description of anchoring as "people make estimates by starting from an initial value that is adjusted to yield the answer" and the observation that the "adjustments are typically insufficient" (p. 1128). Epley and Gilovich (2001) refined this statement by describing the adjustment process via TOTE units (test, operate, test, exit; pp. 392-393). In this line of reasoning, a respondent tests whether the anchor seems plausible (test), adjusts it by adding or subtracting a number (operate), tests again (test), and stops (exit) as soon as the anchor seems plausible (see Figure 1). This model is the only one that provides the basis for measuring parameters other than the total adjustment (which was unreliable in almost all reported cases) and thus derives more specific hypotheses on personality traits that may moderate the susceptibility to anchoring. Different from all the other models, this model can be formalized, and on this basis, adjustment functions with different parameters (e.g., number of adjustment steps, adjustment step size, speed) can be modeled (see Figure 2). By conceptualizing adjustment in this way, adjustment can be understood as the result of a process that involves multiple parameters. This also means that, even if differences in total adjustment are unreliable, there may be underlying parameters that may be reliable. In other words, it may be less important to know *where* one arrived than to understand *how one got there*.

138

Figure 1

Anchoring and Adjustment in the TOTE Model
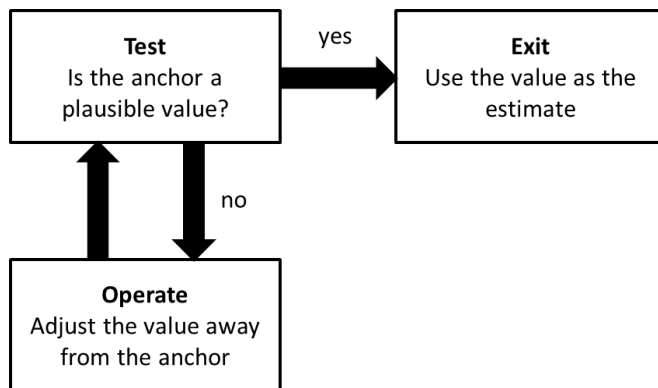
Figure 2

Hypothetical Adjustment Functions from a High Anchor for Two Subjects A and B
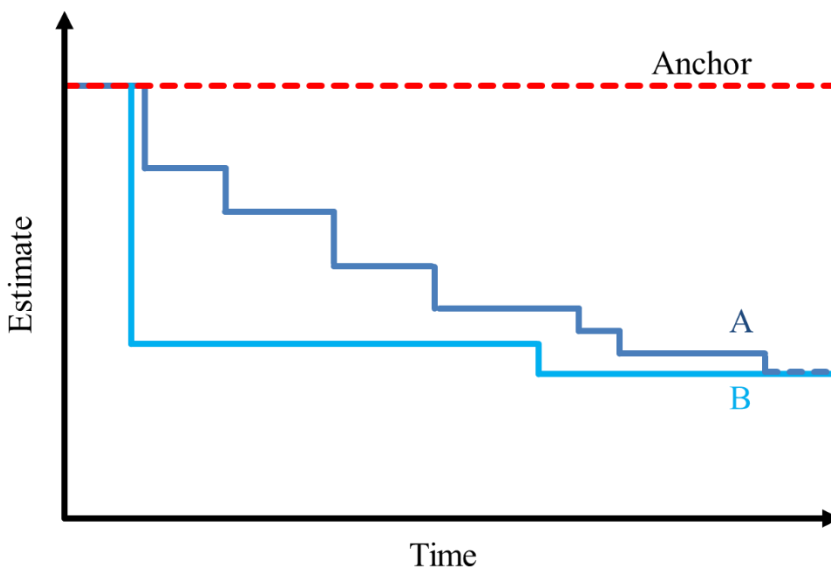
Individual TOTE parameters have already been measured. Frech et al. (2020) introduced a paradigm that required participants to indicate the adjustment steps they made. We used their data (https://osf.io/dq48j/, Study 2) to compute the reliabilities of different parameters and found that the total adjustment ($\alpha = .513$, three items, $N = 146$) and the number of steps ($\alpha = .693$) were reliable, but the average step size ($\alpha = .097$) was not. Moreover, step size decreased in a nonlinear manner such that the initial steps were the largest (for the code and detailed results, see

https://osf.io/yd7c8?view_only=783cf849e6804d678f624e7cda32cd99). In the following, we thus focused on the number of steps as an indicator of adjustment.

What is problematic about Frech et al.'s (2020) task, however, is that only high anchors were used, and participants provided only price estimates—and this is not a typical task in anchoring research. Moreover, the differences were qualitative, that is, the scale had no units, and the distances between the ticks that had been marked were not measured. Instead, participants wrote the numbers next to the ticks. This probably explains why the total adjustment was relatively reliable, too. Still, in the experiment on buying decisions, the total adjustment (i.e., the price estimate) was potentially confounded with the willingness to pay. Other anchoring moderator research has been plagued by similar problems (e.g., Bergman et al., 2010).

To better understand whether the paradigm introduced by Frech et al. (2020) is indeed fit to reliably measure the number of adjustment steps and potentially other adjustment parameters such as adjustment step size, the paradigm has to be adapted and used with classical anchoring effects. That is, adjustment after high *and* low anchors needs to be used for estimates that cannot be confounded with willingness to pay. This is what we tested in the present study. We refer to our version of the approach as the stepwise anchoring paradigm (STAP).

Potential Moderators of the Adjustment Process

We used the TOTE model to derive specific hypotheses on the question of which personality traits predict susceptibility to anchoring. We selected potential moderators by testing for a "common language" between anchoring and the moderators, that is, by considering which personality traits affected the TOTE process: Low cognitive reflection (Frederick, 2005) and high impulsiveness should both be associated with limited considerations and thus with the tendency to quickly affirm the "test" question in the TOTE process, conclude that a number is a

good estimate, and thus stop adjusting. Thus, we argue that both impulsivity and cognitive reflection should predict the number of adjustment steps.

Anchoring and Adjustment as Intuitive and Impulsive Responding

There is little previous research on impulsiveness and anchoring. The closest are reports by Röseler, Schütz, Baumeister, and Starker (2020) and Röseler et al. (2019), who correlated susceptibility to anchoring scores with self-control (Tangney et al., 2004) and found no effect but observed that the total adjustment was not reliable. Considering that intuitive and impulsive answers should lead to an earlier exit from the TOTE process (i.e., "yes, the proposed value is a good answer"), we argue that impulsivity should lead to a smaller number of adjustment steps. The corresponding argument can be made with respect to cognitive reflection: The easiest response to anchoring questions is usually the anchor or a value that is very close to it. This also holds for intuitive responses to the questions from the cognitive reflection task (CRT; Frederick, 2005), and this is why we would predict a positive relationship between the number of adjustment steps and cognitive reflection. In line with this reasoning, associations with other heuristics and biases and the CRT score have been found (e.g., base rate fallacy, conservatism bias, and endowment effect; Hoppe & Kusterer, 2009). But still, findings on anchoring have been heterogeneous: Whereas Meub et al. (2013) and Welsh et al. (2011) reported effects, there have also been reports of null findings (e.g., Bergman et al., 2010; Oechssler et al., 2009; Welsh et al., 2011). In our opinion, the inconclusiveness of previous findings can primarily be attributed to the problem of the low reliability of anchoring effects. This is why we used a new paradigm and tested the reliability of the number of adjustment steps in the anchoring process as a potential alternative indicator of susceptibility to anchoring.

Method

To test the STAP and to determine whether it could provide reliable anchoring parameters (i.e., number of adjustment steps and adjustment step size), we asked participants to estimate the weights of different animals, and we provided high and low anchors. We asked participants to indicate each adjustment step on a scale ranging from 0 to 100 kg. We informed participants of the necessary direction of adjustment (by presenting them with estimates that were too high or too low and had allegedly been made by previous participants) so that the participants' adjustments would be made in only one direction.

Our hypothesis, method, and plan for data analysis were rigorously preregistered (https://osf.io/5ua98?view_only=783cf849e6804d678f624e7cda32cd99). We report how we determined our sample size, all data exclusions, all manipulations, and all measures (Simmons et al., 2012). Slight deviations from the preregistration had to be made (i.e., the analysis script included errors that were fixed before we finished collecting all the data). All changes in the analysis script have been highlighted and commented on.

Planned Sample and Sensitivity

Data were collected online via social networks. Participants were offered the opportunity to participate in a lottery for a 10€ voucher in an online shop and could opt in for feedback on their performance. To provide feedback, we programmed an R (R Core Team, 2018) Shiny-app (Chang et al., 2020) that displayed the number of correct responses in the CRT task, the susceptibility to anchoring, the number of steps, and the z-scored absolute proportional inaccuracy of the estimated animal weights relative to all participants. On the basis of our capacity, we had planned to recruit at least 100 participants with a goal of recruiting approximately 208 participants from June 5 to June 30, 2020.

Hypothesis on the Reliability of Anchoring

As equivalence testing (Lakens et al., 2018) for coefficient alpha is not implemented in common statistical software, we tested for whether the reliability in our paradigm was *lower* (one-tailed test) than the reliability reported by Frech et al. (2020), that is, $\alpha = .69$, three items, $N = 146$. Given a sample size of $N = [100, 208]$, reliability needed to be $\alpha = [.415, .485]$ for us to detect a significant difference with $\alpha = \beta = 5\%$. This was still much higher than commonly estimated reliabilities of anchoring tasks (for an overview, see Röseler et al., 2019, p. 21). Our expectation was that, despite the high power, we would not find a significant difference between the reliability reported by Frech et al. (2020) and the reliability in the present study, and thus, we would be able to conclude that the number of adjustment steps is indeed a reliable anchoring parameter.

Hypotheses on Personality Moderators

Estimates of the correlations between traits and the number of adjustment steps could not be made because there is no previous research on this issue. We thus conducted sensitivity analyses for the personality trait hypotheses, that is, the correlations between the number of adjustment steps and (a) the cognitive reflection task score and (b) the impulsiveness score, too. The minimum Bravais-Pearson correlations for $N = [100, 208]$, $\alpha = \beta = 5\%$, and one-tailed tests needed to be $r = [.321, .225]$. A necessary condition for this hypothesis to be testable is a nonsignificant finding for the reliability hypothesis, that is, the reliability of the number of adjustment steps must be high.

Measures and Materials

The task we used to measure the anchoring parameters was a computerized, modified version of the paradigm used by Frech et al. (2020). Participants estimated the weights of

animals on a scale ranging from 0 to 100 kg (of which only options 1 to 99 kg were selectable due to technical restrictions[10]). For each animal, the anchor that was presented was an estimate that had allegedly been provided by a previous participant and was too high (or too low). Thus, the direction of necessary adjustment was known, which addressed the concern that failing to consider the direction of adjustment may obfuscate potential moderation effects (Cheek & Norem, 2018).

We asked the participants to adjust their estimates by starting from the provided anchor and indicating all steps by clicking on the respective points of the scale. The instructions were adapted from the procedure used by Frech et al. (2020). The respective numeric values were automatically displayed when someone moved the mouse over the scale. A translated example of one item from the anchoring task and the scale are presented in Figure 3. We told participants that adjustment was possible in one direction only: When the anchor was high, we told participants that we would use their lowest estimate as their final estimate, and when the anchor was low, we told them that their highest estimate would be used as their final estimate. To test whether participants had understood the task, we presented them with a sample estimate similar to the one in Figure 3 and asked them to identify the final estimate. Participants were allowed to proceed only if they provided the correct answer.

The 12 items consisted of six pairs of animals that have similar weights. Anchors were chosen proportionally and symmetrically around the true values and in the range of 1 to 99 kg. For each of the pairs of animals, participants were given one high and one low anchor (see Figure 3 for an example with a high anchor). Thus, we were able to compute both within-

---

[10] Due to a technical error, the value of 57 kg was coded as 5 so that we thereby did not know whether a value of 5 represented an actual value of 5 kg or 57 kg. For the items whose low and high anchors were lower and higher than 57, respectively, we checked whether a value of 5 was coded, and we found one case. As a value of 8 was checked, too, we made no changes.

subjects anchoring strength and between-subjects anchoring strength. Participants were randomly assigned to one of two groups and presented with six high and six low anchors. Table 1 provides an overview of all items (animals), actual weights, anchors, and the anchors presented in the two groups.

Table 1

Overview of Item Pairs, True Weights, and Anchors in the Anchoring Task

Pair	Animal	True weight	High anchor	Low anchor
1	Rabbit	4.5	8	1
1	Cat	4.25	8	1
2	Fox	8	14	2
2	Racoon	8	14	2
3	Badger	15	26	4
3	Pelican	15	26	4
4	Beaver	25	40	10
4	Deer	25	40	10
5	Wolf	50	75	25
5	Goat	50	75	25
6	Hyena	60	84	36
6	Warthog	60	84	36

Note. Participants in Group 1 were given the anchors that are highlighted in grey.

Figure 3

Sample Item with Three Highlighted Estimates

The estimate of the person before you was **too high**. What would you estimate?

Estimate of the person before you: **Warthog, 84 kg**

Please indicate **all your mental steps** on the scale. The value that is the farthest from the proposal will be your final estimate.

Note. Light Blue bars indicate estimates. Light Red Bars Indicate 10s. The final estimate is 72.

Cognitive reflection is usually measured with the CRT (Frederick, 2005). However, the reliability of this task is low, too. Multiple versions of the CRT have been proposed (Primi et al., 2016; Thomson & Oppenheimer, 2016; Toplak et al., 2013). In order to achieve high reliability, we used a combination of all of them. An overview of all items, their sources, and the intuitive and correct answers is provided online (https://osf.io/ujyab/?view_only=783cf849e6804d678f624e7cda32cd99). To ascertain whether the participants were familiar with the items and whether this affected the item difficulties, we asked the participants to check all the items (not the solutions) that they had already known before the experiment.

Impulsiveness was measured with the Barrat Impulsiveness Scale (BIS-11), which consists of 30 items, such as "I do things without thinking" (Stanford et al., 2009). We included

an attention check as the 16th item of the BIS-11, which asked participants to check "sometimes."

Procedure

First, participants were greeted, informed about the purpose of the experiment (i.e., to estimate the body weights of animals), and asked to answer demographic questions (age, gender, highest educational level achieved). Then they completed the BIS-11, the attention-check item, and the CRT items. After they were informed about how the anchoring task worked, the participants estimated the animal weights. Finally, we presented the CRT items again (without the solutions) and asked the participants to indicate any questions they had already known the answers to before the study and whether they had looked up the answers. Participants had the option to enter their e-mail address to take part in a lottery and the option to receive a link to the feedback. An overview of the procedure is presented in Figure 4.

Figure 4

Study Procedure

Greeting and Purpose of Experiment
Demographic Data age, gender, educational level
BIS-11 30 Impulsiveness self-report items and attention check
CRT 13 Cognitive reflection task items
Anchoring task 12 items
CRT check Check the questions that you already knew
Lottery participation and feedback

Data Analysis Plan

Power and sensitivity analyses, data processing, data analysis, and the feedback app were conducted with R (R Core Team, 2018) version 3.6.3 and the packages pwr (Champely, 2020), psych (Revelle, 2018), cocron (Diedenhofen, 2016), dmetar (Harrer et al., 2019), meta (Balduzzi et al., 2019), stringr (Wickham, 2018), cocor (Diedenhofen & Musch, 2015), Hmisc (Harrel, Frank, E., 2020), and shiny (Chang et al., 2020). Note that the entire analysis script was preregistered. All changes (e.g., due to coding errors) are highlighted and commented on.

The BIS-11 was computed as the mean of all items after reverse scoring the negatively keyed items. The CRT score was the sum of all correct answers. We allowed slight deviations from the true values in Table 1, such as "five" instead of "5" or "Sahra" instead of "Sarah" (see the analysis script for the exact code, https://osf.io/pvqdk/?view_only=783cf849e6804d678f624e7cda32cd99).

The number of adjustment steps was the number of checks on the scale. We subtracted 1 from the score if the anchor had been checked, too. The final estimate was the lowest value for high anchors and the highest value for low anchors. The total adjustment was the absolute difference between the anchor and the final estimate. The mean adjustment step size was the total adjustment divided by the number of adjustment steps. The within-subjects susceptibility to the anchoring score was computed as the high anchor minus the low anchor. The mean susceptibility to the anchoring score was the mean of the six z-transformed scores from the six pairs. See Figure 5 for a graphical representation of the anchoring parameters on the scale used in the 12 anchoring tasks.

Participants were excluded from the analysis if any of the following preregistered criteria were not met: (a) age was more than 17 years and less than 123 years, (b) completed all BIS-11 items, (c) passed the attention check, (d) showed variability in responses to the BIS-11, (e)

estimate was not the anchor, (f) did not skip any anchoring item, (g) reported that they did not look up the CRT answers.

Figure 5

Anchoring (TOTE) Parameters as Operationalized from the Scale used in the Anchoring Task
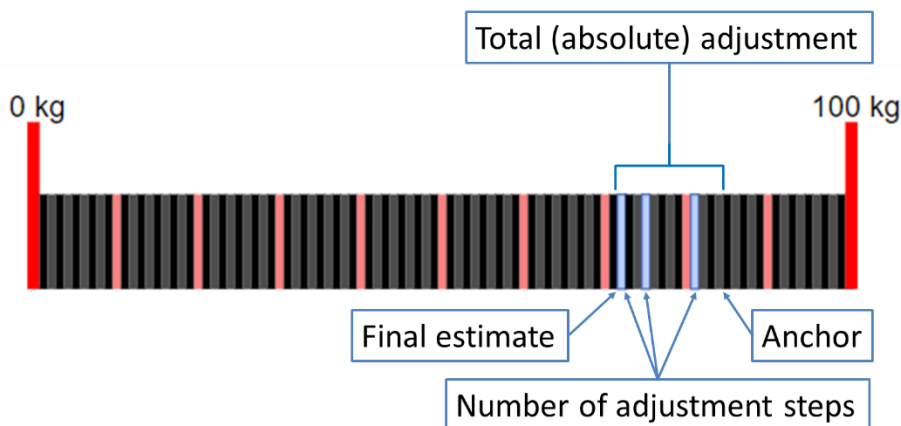
Note. In the example, the anchor was 84 kg (too high). There were three adjustment steps with 72 kg as the final estimate, and the total adjustment was thus 12 kg. The mean adjustment step size was thus 3 kg per step.

Results

If not indicated otherwise (e.g., by the heading "Exploratory Analyses"), all reported results were preregistered. We highlighted changes from the preregistered analysis script in the script and discussed them if they involved aspects that went beyond coding errors.

Achieved Sample and Sensitivity

The achieved sample size was $N = 123$ (82 women; mean age = 28.79 years). A total of 66 participants were assigned to the "grey group" (see Table 1; i.e., rabbit was paired with a high anchor and cat with a low anchor). A sensitivity analysis revealed that the largest reliability for which we could have found a difference from the one reported by Frech et al. (2020) with $\alpha = \beta = .05$ was $\alpha = .505$. Given these parameters, the smallest detectable correlation between the number of adjustment steps and the personality traits was $r = .290$ (one-tailed test).

Reliability of Measures

The reliability of the CRT items was $\alpha = .802$ (13 items, $N = 123$). The BIS-11 impulsiveness scale reliability was $\alpha = .788$ (30 items, $N = 123$).

Manipulation Checks

We determined the strength of the anchoring effects between subjects (based on the two groups in Table 1) and within subjects (based on the estimates for the pairs of animals with equal weights but high and low anchors). We computed Cohen's d for all 12 items and conducted a random-effects meta-analysis to determine whether the overall effect was significantly larger than zero, which it was, $t(11) = 10.03$, $p < .001$, $d = 0.977$, 95% CI $= [0.763, 1.192]$. A forest plot of all anchoring effects is displayed in Figure 6. Within-subjects anchoring effects were calculated for all six pairs of items (all ps $< .001$) with an overall large and significant anchoring effect, $t(5) = 5.10$, $p = .004$, $dz = 0.749$, 95% CI $= [0.371, 1.126]$. In other words, there were anchoring effects on the individual level in 70% of all cases.

Figure 6

Forest Plot of Between-Subjects Anchoring Effects for All Items

No.	Item	Cohen's d	95%-CI
1	rabbit		[0.23; 0.95]
2	cat		[0.72; 1.44]
3	fox		[1.11; 1.93]
4	racoon		[0.02; 0.65]
5	badger		[0.56; 1.28]
6	pelican		[0.36; 1.06]
7	beaver		[0.50; 1.19]
8	deer		[1.00; 1.76]
9	wolf		[0.65; 1.36]
10	goat		[0.78; 1.57]
11	hyena		[0.69; 1.41]
12	warthog		[0.75; 1.54]

Random effects model [0.76; 1.18]
Prediction interval [0.35; 1.59]
Heterogeneity: $I^2 = 70\%$, $\tau^2 = 0.0678$, $p < 0.01$
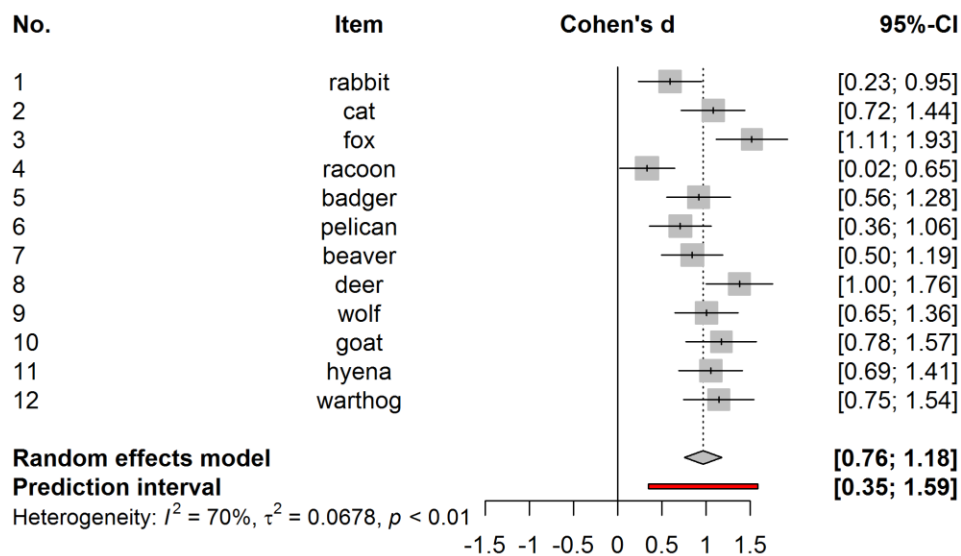
Figure 7

Forest Plot of Within-Subjects Anchoring Effects for All Items

No.	Item	Cohen's dz	95%-CI
1	rabbit-cat		[0.17; 0.51]
2	fox-racoon		[0.78; 1.19]
3	badger-pelican		[0.30; 0.66]
4	beaver-deer		[0.83; 1.24]
5	wolf-goat		[0.33; 0.69]
6	hyena-warthog		[0.96; 1.39]

Random effects model [0.38; 1.12]
Prediction interval [-0.18; 1.68]
Heterogeneity: $I^2 = 92\%$, $\tau^2 = 0.0926$, $p < 0.01$
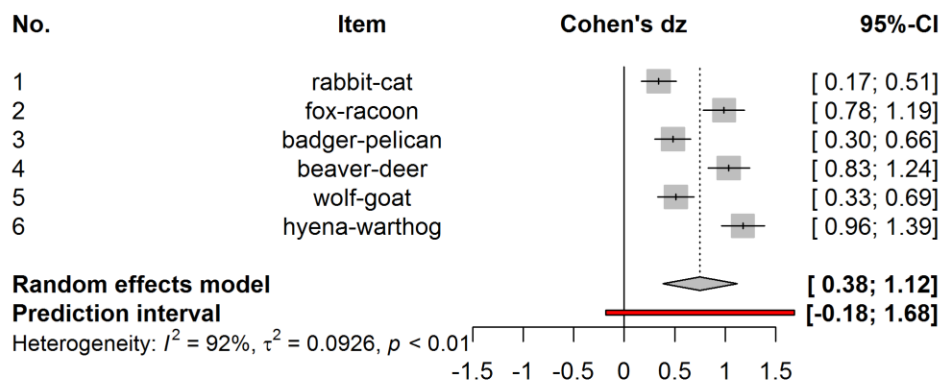
Hypothesis Tests

The reliability of the number of adjustment steps was acceptable, $\alpha = .749$, 12 items, $N = 123$, and was not significantly smaller than Frech et al.'s (2020) reliability ($\alpha = .69$, three items, $N = 146$), $F(122, 145) = 1.24$, $p = .889$ (one-tailed test). An exploratory analysis revealed that the reliability in our task was not significantly different (i.e., not higher) either, $F(122, 145) = 1.24$, $p = .221$ (two-tailed test, exploratory analysis).

For the correlations with personality traits, contrary to our prediction, the number of adjustment steps was not positively correlated with cognitive reflection, $r(121) = -.218$, $p = .992$ (one-tailed) or negatively correlated with impulsiveness, $r(121) = .111$, $p = .889$ (one-tailed). Note that the directions of both correlations were opposite the hypothesized directions.

Exploratory Analyses

The reliabilities of the other anchoring parameters were lower than the reliability of the number of adjustment steps. An overview of the Cronbach's alphas for the other parameters as well as the measures reported above is provided in Table 2. Despite their low reliability, the correlations between the other anchoring parameters and the personality traits were possible but were not observed. Correlations involving all the variables from the experiment are presented in Table 3.

Table 2

Reliabilities of the Constructs Measured in the Study

Parameter	Number of items	Average r	Coefficient α	Construct
Absolute adjustment[a]	12	.052	.399	Anchoring
Number of adjustment steps[a]	12	.213	.749	Anchoring
Adjustment step size[a]	12	.126	.633	Anchoring
WS difference score[a]	6	-.086	-.915	Anchoring
Anchor index	12	.012	-.260	Anchoring
Weight estimates	12	.126	.634	Weight estimates
CRT 13[a]	13	.237	.802	Cognitive reflection
CRT 3	3	.388	.656	Cognitive reflection
BIS-11[a]	30	.110	.788	Impulsiveness

Note. $N = 123$. WS = within subjects. Anchor index is absolute adjustment divided by the anchor.

[a]Preregistered analysis.

Table 3

Correlation Matrix of Anchoring Parameters and Personality Traits

No.	Variable	1	2	3	4	5
1	Adjustment steps	*.749*	-.499*	-.750*	.111	-.218
2	Absolute adjustment	-.499*	*.399*	.932*	.214	.012
3	Step size	-.750*	.932*	*.633*	.120	.098
4	BIS-11	.111	.214	.120	*.788*	-.154
5	CRT 13	-.218	.012	.098	-.154	*.802*

Note. Values in italics represent Cronbach's alphas.

*p < .001.

Besides comparing Cronbach's alphas, average correlations could be compared, too. After all, Cronbach's alpha increases with the number of items, but the average correlation does not. The average interitem correlation for the number of adjustment steps in our study ($r = .213$, 12 items, $N = 123$) was relatively high in comparison with other reported values (for an overview, see Röseler et al., 2019) and was significantly smaller than in Study 2 in Frech et al. (2020; from whose data we had computed $r = .431$, three items, $N = 146$), $Z = -1.97$, $p = .049$ (two-tailed, see Table 2 for descriptives). Given that this was one of many exploratory analyses and alpha inflation was an issue, we advise readers to interpret this result with caution.

Although we used almost the same instructions as Frech et al. (2020) did, the distribution of the number of steps was different in our experiment than in the study by Frech et al. (2020). For example, the mean number of steps in our study was $M = 2.11$ ($SD = 0.37$, $N = 123$) and thus smaller than in the study by Frech et al. (2020, $M = 3.80$, $SD = 1.30$, $N = 146$), $t(267) = 13.95$, $p < .001$, $d = 1.768$. A possible reason is that people are more familiar with stepwise estimates for price negotiations than when estimating weights.

As we had asked participants whether they had already known the CRT items (not the answers to them), the exclusion of items that are well-known had the potential to change the scale and thus the correlation with the number of adjustment steps. The proportion of people who knew the items was extremely high ($p_{min} = .16$), and the bat-and-ball item was familiar to 45% of the participants. Surprisingly, however, the number of known items was not correlated with the number of correct answers, $r(121) = -.014$, $p = .877$ (two-tailed). Direct feedback from some participants suggested that some of them misunderstood the question and reported whether they *remembered the listed items from earlier in the experiment* instead of whether they had known the items.

Discussion

Although personality moderator research represents a major part of the anchoring literature from the past 13 years (e.g., McElroy & Dowd, 2007), nothing has changed since Furnham and Boo (2011) stated that "researchers have failed to identify any … trait variables that have a systematic and explicable effect on anchored decisions" (p. 40). This is most likely because anchoring scores have demonstrated low reliability (Röseler et al., 2019). We solved this problem by adapting a paradigm by Frech et al. (2020), which was based on the insufficient adjustment model of anchoring (e.g., Epley & Gilovich, 2001) and required participants to indicate all of their adjustment steps: First, we reanalyzed the data provided by Frech et al. (2020) and found that the *number of adjustment steps* was a much more reliable indicator of susceptibility to anchoring than the traditionally used absolute total adjustment from the anchor based on the data provided by Frech et al. (2020).

Second, we adapted the paradigm that Frech et al. (2020) had used with buying decisions and high anchors and used it with traditional anchoring items (weight estimation), a larger number of items (12 instead of three), and both high and low anchors. Furthermore, we used a repeated-measures manipulation of the anchor and were thus able to estimate within-subjects anchoring effects. An analysis of four different anchoring parameters (i.e., total absolute adjustment, number of adjustment steps, average adjustment step size, and within-subjects difference scores) corroborated the hypothesis that the number of adjustment steps is the most reliable parameter in anchoring effects. The results are in line with those from Frech et al. (2020) and open the door to methodologically sound anchoring moderator research.

On the basis of the TOTE model of anchoring, we hypothesized that an early exit from TOTE loops (i.e., a small number of adjustment steps) would be related to low cognitive

reflection or high impulsiveness. However, we found no evidence for either of these hypotheses. This leaves open the question of what psychological constructs may be associated with the number of adjustment steps or other adjustment parameters in anchoring.

Limitations and Future Directions

For future research to adopt the STAP, we advise researchers to include even more items to achieve even higher reliability. In previous research that used the total adjustment as the dependent variable, this aspect was not considered, and the use of repeated measures did not help because the dependent variables were not correlated. If, however, the number of adjustment steps were to be used, approximately 37 items would be needed to achieve a reliability beyond $\alpha = .90$. This would also reduce concerns about power. Given that the minimum detectable correlation in our study was $r = .290$ ($\alpha = \beta = .05$, $N = 123$, one-tailed), a larger sample size would be especially desirable when conducting personality moderator research (e.g., Funder & Ozer, 2019).

Furthermore, animal weight estimation and price estimation are limited by the number of animals or products. Considering the development of a diagnostic tool for assessing susceptibility to anchoring, future paradigms should use stimuli that are unlimited and thus more practical for use in repeated-measures designs, for example, estimates of the number of dots in a scatterplot (e.g., Röseler et al., 2019). Although a hyena and a warthog weigh about the same, their estimated weights may differ. This would not be the case for the number of points in a scatterplot: Most people would not realize if they were shown 50 randomly distributed points twice.

In our study, we did not discuss whether the anchor should be random or informative. Framing our anchors as estimates from a previous participant rendered the anchors informative

rather than random. Although research has shown that random anchors work, too (e.g., Englich et al., 2006), some have found that random anchors have smaller effects than informative anchors framed as suggestions (e.g., Röseler et al., 2019, pp. 13-14), and some studies have even observed no effect at all in the case of random anchors (e.g., Schweickart et al., 2020). Future research needs to determine the degree to which random anchors lead to anchoring effects and why effects for informative anchors may be different. The STAP provides a useful tool for addressing this question. For example, adjustment from random anchors could involve more or larger adjustment steps.

Sources of variation of reliability between studies are still unknown and need to be investigated. For example, why was the total adjustment more reliable in our study than in other studies? The most problematic case would be if the STAP altered the anchoring process by "dictating" to participants that they needed to make adjustment steps in the form of a conversational norm or demand characteristics (Grice, 1982). This concern can partly be ruled out because, for example, the effect of precise (vs. round) anchors on anchoring effect strength is similar in this paradigm (Frech et al., 2020) as it is in the traditional one (Janiszewski & Uy, 2008; Loschelder et al., 2014; Loschelder et al., 2017; Mason et al., 2013). To be even better able to rule out the concern that the number of adjustment steps represents only a methodological artifact, we suggest that different paradigms need to be developed to assess TOTE parameters such as the number of adjustment steps.

The most pressing problem in anchoring and personality trait research is the lack of any substantial and replicable correlation between anchoring parameters and personality traits. Although the present study provides a reliable measure of susceptibility to anchoring, we do not yet know about the validity or more specifically the construct validity of this measure. We also

do not know why the *number* of adjustment steps but not the *size* of the adjustment steps is reliable. In the worst case, the number of adjustment steps might not even represent the loops involved in the TOTE process but something else. One way to establish the construct validity of the adjustment steps is to manipulate the number of steps—analogous to Frech et al. (2020) who successfully manipulated the average step size by providing exact anchors (i.e., numbers with fewer zeros in them). Some participants gave us the personal feedback that they were worried because they made only a single adjustment step by just guessing a final value immediately. It is undisputed (Epley & Gilovich, 2001) and in line with Frech et al.'s (2020) and our data that there are people who make multiple adjustment steps. However, it is possible that there are also some people who make only a single adjustment step.

There remain a wealth of parameters that can be integrated in the TOTE model, such as the time needed to make adjustment steps, the (non)linearity of adjustment, or whether all steps are made in the same direction. In fact, some participants asked us why the adjustments could only be made in one direction, which suggests that the applied model may be too simple in that it does not permit the second step to be closer to the anchor than the first step. Further analyses of Frech et al.'s (2020) data showed that there were a few individuals who not only adjusted *away* from the anchor but in a later step also adjusted backwards *toward* the anchor. The paradigm used in this study is only one way to measure these parameters. Other models that incorporate parameters differently or models that use more parameters might be even more suitable for the measurement of anchoring and adjustment parameters. Future research needs to find the optimal number of parameters while still maintaining a certain degree of parsimony. Some possible TOTE models of anchoring are summarized in Figure 8 and Table 4. The model we used can be

characterized by a fixed direction of adjustment, variable step size, unknown time between steps, and discrete steps.

Figure 8

Different TOTE Models Allow for Different Possible Adjustment Functions
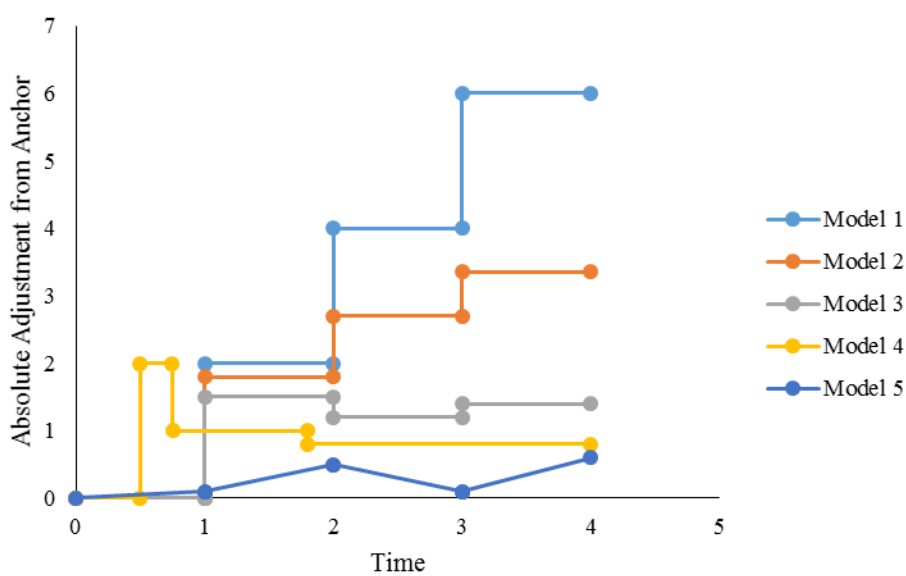

Table 4

Overview of Model Parameters from Figure 8

Model	Direction of adjustment	Step size	Time between steps	Nature of steps
1	Fixed	Fixed	Fixed	Discrete
2	Fixed	Decreasing	Fixed	Discrete
3	Variable	Variable	Fixed	Discrete
4	Variable	Variable	Variable	Discrete
5	Variable	Variable	Variable	Continuous

Conclusion

Individual differences in anchoring have been investigated for more than a decade. Still, researchers have overlooked a crucial but necessary condition for valid results, namely, reliability. In this paper, we tested the STAP and can state that it is currently the first and only way to assess anchoring and adjustment reliably by breaking down the adjustment process into

the number of adjustment steps and the sizes of these steps. We hope that this progress may lay the groundwork to resolve the inconsistencies that resulted from a decade of research on individual differences in anchoring to finally yield explicable and consistent findings on individual susceptibility to anchoring.

Open Practices Statement

All study materials, data, the analysis script (https://osf.io/fjxmw/?view_only=783cf849e6804d678f624e7cda32cd99), and the preregistration (https://osf.io/5ua98/?view_only=783cf849e6804d678f624e7cda32cd99) are available online. Note that minor changes due to coding errors were made to the analysis script after the preregistration. We highlighted and explained all of these changes in the analysis script.

Author Contributions

LR planned and programmed the study and the analysis script and created the preregistration and the first draft of the manuscript. IKD, KF, YH, JCH, ELEI, AHK, SK, DS, and AZ created the study materials and recruited participants. JJR checked and revised the analysis script. All authors revised and approved the manuscript for publication.

References

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The Sticky Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues. *Journal of Behavioral Decision Making*, *87*(1), 23. https://doi.org/10.1002/bdm.2022

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

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Champely, S. (2020). *pwr: Basic Functions for Power Analysis. R package version 1.3-0* [Computer software]. https://CRAN.R-project.org/package=pwr

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2020). *shiny: Web Application Framework for R. R package version 1.4.0.2* [Computer software]. https://CRAN.R-project.org/package=shiny

Cheek, N. N., & Norem, J. K. (2018). On Moderator Detection in Anchoring Research: Implications of Ignoring Estimate Direction. *Collabra: Psychology*, *4*(1), 12. https://doi.org/10.1525/collabra.125

Cheek, N. N., & Norem, J. K. (2019). Are Big Five Traits and Facets Associated With Anchoring Susceptibility? *Social Psychological and Personality Science*, *92*, 194855061983700. https://doi.org/10.1177/1948550619837001

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, *21*(3), 241–251. https://doi.org/10.1002/bdm.586

Diedenhofen, B. (2016). *cocron: Statistical Comparisons of Two or more Alpha Coefficients* [Computer software]. http://comparingcronbachalphas.org

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One*, *10*(3), e0121945. https://doi.org/10.1371/journal.pone.0121945

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

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a Within-Subject, Repeated-Measures Ego-Depletion Paradigm. *Social Psychology*, *49*(5), 271–286. https://doi.org/10.1027/1864-9335/a000348

Frech, M.-L., Loschelder, D. D., & Friese, M. (2020). How Attribution-of-Competence and Scale-Granularity Explain the Anchor Precision Effect in Negotiations and Estimations. *Social Cognition*, *38*(1), 40–61. https://doi.org/10.1521/soco.2020.38.1.40

Frederick, S. W. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology. General*, *141*(1), 124–133. https://doi.org/10.1037/a0024006

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Furnham, A., Boo, H. C., & McClelland, A. (2012). Individual Differences and the Susceptibility to the Influence of Anchoring Cues. *Journal of Individual Differences*, *33*(2), 89–93. https://doi.org/10.1027/1614-0001/a000076

Grice, H. P. (1982). *Speech acts* (5. ed.). *Syntax and semantics: Vol. 3*. Academic Press.

Harrel, Frank, E. (2020). *Hmisc: Harrell Miscellaneous. R package version 4.4-0* [Computer software]. https://CRAN.R-project.org/package=Hmisc

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'* (Version 0.0.9000) [Computer software]. http://dmetar.protectlab.org

Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., Urlichich, D., & Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology. General.* Advance online publication. https://doi.org/10.1037/xge0000644

Hoppe, E. I., & Kusterer, D. J. (2009). Behavioral Biases and Cognitive Reflection. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.1488752

Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science*, *19*(2), 121–127. https://doi.org/10.1111/j.1467-9280.2008.02057.x

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., Babalola, M. T., Bahník, Š., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Bocian, K., Brandt, M., Busching, R., Cai, H., Cambier, F., . . . Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. Advance online publication. https://doi.org/10.31234/osf.io/9654g

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Loschelder, D. D., Friese, M., & Trötschel, R. (2017). How and why precise anchors distinctly affect anchor recipients and senders. *Journal of Experimental Social Psychology*, *70*, 164–176. https://doi.org/10.1016/j.jesp.2016.11.001

Loschelder, D. D., Stuppi, J., & Trötschel, R. (2014). "€14,875?!": Precision Boosts the Anchoring Potency of First Offers. *Social Psychological and Personality Science*, *5*(4), 491–499. https://doi.org/10.1177/1948550613499942

Mason, M. F., Lee, A. J., Wiley, E. A., & Ames, D. R. (2013). Precise offers are potent anchors: Conciliatory counteroffers and attributions of knowledge in negotiations. *Journal of Experimental Social Psychology*, *49*(4), 759–763. https://doi.org/10.1016/j.jesp.2013.02.012

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making*, *2*(1), 48–53.

Meub, L., Proeger, T., & Bizer, K. (2013). Anchoring: A Valid Explanation for Biased Forecasts When Rational Predictions are Easily Accessible and Well Incentivized? *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2307619

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., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes*, *98*(2), 133–143. https://doi.org/10.1016/j.obhdp.2004.12.002

Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgemental anchoring. *European Review of Social Psychology*, *10*(1), 135–167. https://doi.org/10.1080/14792779943000044

Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, *35*, 136–164. https://doi.org/10.1037/e413782005-562

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), 147–152. https://doi.org/10.1016/j.jebo.2009.04.018

Petty, R. E., DeMarree, K. G., Briñol, P., Horcajo, J., & Strathman, A. J. (2008). Need for cognition can magnify or attenuate priming effects in social judgment. *Personality & Social Psychology Bulletin*, *34*(7), 900–912. https://doi.org/10.1177/0146167208316692

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), 453–469. https://doi.org/10.1002/bdm.1883

R Core Team. (2018). *R* [Computer software]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research, Version = 1.8.3* [Computer software]. Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych

Röseler, L., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce judgment adjustment for both internally and externally generated anchors? *Journal of Experimental Social Psychology*, *87*, 103942. https://doi.org/10.1016/j.jesp.2019.103942

Röseler, L., Schütz, A., Blank, P. A., Dück, M., Fels, S., Kupfer, J., Scheelje, L., & Seida, C. (2020). No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered, and Failed Replication Attempts. Manuscript submitted for publication.

Röseler, L., Schütz, A., & Starker, U. (2019). Cognitive Ability does not and cannot Correlate with Susceptibility to Anchoring Effects. Advance online publication. https://doi.org/10.31234/osf.io/bnsx2

Schweickart, O., Tam, C., & Brown, N. R. (2020). When "bad" is good: How evaluative judgments eliminate the standard anchoring effect. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000209

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2160588

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, *47*(5), 385–395. https://doi.org/10.1016/j.paid.2009.04.008

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Teovanović, P. (2019). Individual Differences in Anchoring Effect: Evidence for the Role of Insufficient Adjustment. *Europe's Journal of Psychology*, *15*(1), 8–24. https://doi.org/10.5964/ejop.v15i1.1691

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113. http://journal.sjdm.org/15/151029/jdm151029.html

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

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

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W.B. G. (2001). Implications of Attitude Change Theories for Numerical Anchoring: Anchor Plausibility and the Limits of Anchor Effectiveness. *Journal of Experimental Social Psychology*, *37*(1), 62–69. https://doi.org/10.1006/jesp.2000.1431

Welsh, M. B., Delfabbro, P. H., Burns, N. R., & Begg, S. H. (2011). Individual differences in anchoring: Numerical ability, education and experience. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33).

Wickham, H. (2018). *stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.0* [Computer software]. https://CRAN.R-project.org/package=stringr

Wood, S., Poynor, C., & Chartrand, T. (2011). Individual Susceptibility to Priming Effects. *ACR North American Advances*.

Yarkoni, T. (2020, May 6). *Induction is not optional (if you're using inferential statistics): reply to Lakens [Blog post]*. https://www.talyarkoni.org/blog/2020/05/06/induction-is-not-optional-if-youre-using-inferential-statistics-reply-to-lakens/

Article 5: Downsides of High Self-Control

The following manuscript is the accepted version of a manuscript that has been accepted at Europe's Journal of Psychology and is to be published in 2021.

Röseler, L., Ebert, J., Schütz, A., & Baumeister, R., F. (2021). The Upsides and Downsides of High Self-Control: Evidence for Effects of Similarity and Situation Dependency. *Europe's Journal of Psychology.*

The Upsides and Downsides of High Self-Control: Evidence for Effects of Similarity and

Situation Dependency

Lukas Röseler[1, 2]

Jacqueline Ebert[1]

Astrid Schütz[1]

Roy F. Baumeister[3]

[1]Department of Personality Psychology and Psychological Assessment, University of Bamberg,

Bamberg, Germany

[2]Department of Economics, Harz University of Applied Sciences, Wernigerode, Germany

[3]School of Psychology, University of Queensland, Brisbane, Australia

Corresponding Author
Lukas Röseler
lukas.roeseler@uni-bamberg.de
Department of Personality Psychology and Psychological Assessment
Markusplatz 3, 96047 Bamberg, Germany

Abstract

High trait self-control is generally depicted as favorable. We investigated whether this holds for social perception. Using vignettes, we tested whether a person with high self-control is (a) preferred as a partner for all or only certain social situations, (b) perceived as less likeable than a person with low self-control, (c) liked more if the person is female and the behavior thus fits the sex-stereotype, and (d) perceived differently from a person with low self-control with respect to a wide range of adjectives used to describe personality. Competing theories are presented for each area. Results indicate that although high self-control is associated with a wide range of socially desirable traits, choice of partners (a) depends on the type of situation in which the interaction will occur, (b) depends on the similarity between the respondent and the partner, (c) does not depend on a stereotype match, and (d) does not depend or depends only to a small degree on the partner's high self-control. The perception of individuals with high self-control is thus variable and situationally contingent, and more than a single theory is needed to explain it.

Keywords: self-control; social perception; attraction

The Downsides of High Self-Control: Evidence for Effects of Similarity and Situation

Dependency

Self-control, which is defined as people's capacity to alter their own responses and bring

them in line with things such as social expectations, is generally considered a socially desirable

characteristic (e.g., Baumeister, Vohs, & Tice, 2007). Indeed, even the measurement of self-

control invokes the assumption of benefits. One self-control item is "I do certain things that are

bad for me…" (reverse scored; Tangney, Baumeister, & Boone, 2004). In fact, it has been shown

that self-control is advantageous in many areas of life, such as subjective well-being, binge-

eating, alcohol use, grades in school, commitment in a relationship (Ridder, Lensvelt-Mulders,

Finkenauer, Stok, & Baumeister, 2012), malevolent and fractious intentions, outward directed

aggression (Tangney et al., 2004), or selfishness and helpfulness (Dewall, Baumeister, Gailliot,

& Maner, 2008). In a way, self-control implies restraining bodily urges that include short-term

pleasure. Thus, there is an assumption that people will act in selfish and reward- or lust-oriented

ways (e.g., the selfishness hypothesis of ego depletion; Banker, Ainsworth, Baumeister, Ariely,

& Vohs, 2017) when no superordinate entity is present to exert control over them (e.g., Hobbes

& Brooke, 1651/2017). Self-control would thus be fundamentally and pervasively beneficial. We

refer to this as the positivity hypothesis, that is, the idea that high self-control is either beneficial

or irrelevant. Like intelligence, self-control can be compared to a tool. Tools are good to have.

To be sure, tools can be used for bad ends, but that does not reflect any fault of the tools

themselves.

Past research has occasionally looked for disadvantages of high self-control but without

much success. For example, no support was found for the hypothesis that too much self-control

(i.e., a curvilinear relationship between self-control and other variables, so that intermediate

levels of self-control produce the best results) could contribute to compulsions and eating disorders. Instead, Tangney et al. (2004) found a linear negative relationship between self-control and the occurrence of impulse control problems (e.g., binge eating) and self-reports of psychopathological symptoms (e.g., obsessive-compulsive patterns, anxiety). About the worst that can be said about high self-control is that individuals with high self-control tend to be confronted with higher expectations and are therefore more likely to suffer from exhaustion (Koval, vanDellen, Fitzsimons, & Ranby, 2015) and that high functional impulsivity (a variation of low self-control) can be advantageous when there is a need to process information quickly (Dickman & Meyer, 1988).

Even less attention has been devoted to whether interpersonal problems are related to high self-control, even though striving and conformity are not necessarily reflected in positive relationships. For example, students who get good grades in school are more likely to be labeled nerds, which is a slang term that disparages scholarly achievement (Pelkner & Boehnke, 2003; Rentzsch, Schütz, & Schröder-Abé, 2011). Even so, such stigmatizing seems to reflect anti-intellectual resentment more than a genuine social disadvantage of high self-control. Much work has confirmed that people with high self-control are more popular and have better relationships than other people (e.g., Maszk, Eisenberg, & Guthrie, 1999; Vohs, Finkenauer, & Baumeister, 2011).

Derivation of Hypotheses

In the present study, we took a social perception perspective to investigate whether there may be interpersonal downsides of high trait self-control. To cover a broad area, we included (a) preferences for partners with high versus low self-control in various social situations, (b) liking and stereotypes of people with high versus low self-control, and (c) perceivers' personality traits.

Downsides of High Self-Control in Social Interactions

Assuming that a person with low self-control agrees with the statement "I do certain things that are bad for me, if they are fun" (Tangney et al., 2004), that person might be considered good company if there is a lively party with lots of food and drink. Such activities are common: Among Germans between 18 and 25 years of age, 33.4% report engaging in the regular consumption of alcohol (BZgA, 2019). Thus, being self-controlled may interfere with being perceived as good company during leisure time and at parties.

Of course, everyday situations differ with regard to the kinds of behavior that are expected. Taxonomies such as the one by Rauthmann et al. (2014) have distinguished types of situations such as sociality and duty. Duty situations typically feature aspects that are not fun but need to be done (e.g., doing homework or cleaning the house). Socializing situations are often hedonic (e.g., going to a party or cooking a meal together). We hypothesized that whether people would prefer the company of a person with high or low self-control would depend on the situation they are in. Note that Lösch and Rentzsch (2018) found that students like to *work* with partners with high conscientiousness (which overlaps with self-control), but during leisure time, people with high conscientiousness are not chosen over people with low conscientiousness.

We propose three competing hypotheses with regard to the interaction of self-control and situations. Our thinking invoked Mischel's (1968) distinction between strong and weak situations (see also Cooper & Withey, 2009). In this understanding, strong situations leave little or no room for different behaviors (e.g., everyone stops at a red light when driving a car) whereas weak situations leave much room (e.g., people behave differently during a private party). The hypotheses thus differ with respect to how strongly the situation (duty vs. socializing) impacts preferences for an interaction partner with high vs. low self-control.

1) Strong situation hypothesis: In situations that call for duty, people prefer to be with a person with high self-control rather than a person with low self-control. In socializing situations, people prefer to be with a person with low self-control rather than a person with high self-control (high situational strength for duty *and* socializing).

2) Weak situation hypothesis: In duty situations, people prefer to be with a person with high self-control rather than a person with low self-control. In socializing situations, people have no preference (high situational strength for duty and low situational strength for socializing).

3) Positivity hypothesis: In any type of situation, people prefer to be with a person with high self-control rather than a person with low self-control (no situational strength).

Perceivers' Traits in Relation to Targets' Traits

One of the main predictors of whether a person likes another person is similarity (Byrne & Nelson, 1965). The so-called similarity hypothesis has received a lot of support with regard to personality traits such as optimism in romantic relationships (Böhm, Schütz, Rentzsch, Körner, & Funke, 2010), personality and physical attractiveness in roommate relationships (Carli, Ganley, & Pierce-Otay, 1991), humor style in married couples (Hahn & Campbell, 2016), gender-role self-concept and selection of dating partners (Morell, Twillman, & Sullaway, 1989), socially undesirable traits such as Machiavellianism (Ináncsi, Láng, & Bereczkei, 2016), and same-sex friendships regardless of duration or closeness (Morry, 2005). The meta-analytic effect of interpersonal attraction and perceived similarity was estimated to be $r = .39$ (Montoya, Horton, & Kirchner, 2008). However, hypotheses that propose effects that go in the opposite direction have also been formulated: According to the complementarity hypothesis, which is based on the circumplex theory (Kiesler, 1983), relationships work best when the two partners have different scores with regard to dominance but not with regard to affiliation. This finding

was supported with respect to dependent variables such as liking in the context of dominant and submissive nonverbal behaviors (Dryer & Horowitz, 1997; Tiedens & Fragale, 2003) or relationship quality and facial attractiveness (Asendorpf, Penke, & Back, 2011, p. 25, Hypothesis 3b), but it was also challenged by null findings (Asendorpf et al., 2011, p. 25). Boman (2016) found that friends' levels of self-control are largely unrelated in both attitudinal and behavioral measures. With regard to self-control, Vohs et al. (2011) found that romantic couples seemed to pair up in a complementary fashion, such as that dating partners were less similar to each other than they would be by chance (p. 143). The effect was weaker (but still significant) among married couples and absent among same-sex friends, suggesting that the attraction was primarily sexual. Neither similarity nor complementarity predicted relationship satisfaction, which was simply (and positively) correlated with the total of both partners' self-control, consistent with the positivity hypothesis.

Third, there is the compensation hypothesis, according to which qualities of one partner contribute to the other's satisfaction and may even partially compensate for the partner's lack of these traits (Schröder-Abé & Schütz, 2011). Thus, individuals with low self-control may prefer individuals with high self-control to compensate for their own lack of self-control. Shea, Davisson, and Fitzsimons (2013) provided evidence for such effects with respect to self-control and confirmed the compensation hypothesis when they tested it against the other theories for both trait and state self-control. However, compensation worked for low self-controlled individuals only. That is, low self-controlled individuals preferred high self-controlled partners but high self-controlled partners had no preference.

On the basis of these models, we hypothesized that liking would be a function of people's own trait self-control and the trait self-control described in the vignettes, a tendency that has

been observed for optimism (Böhm et al., 2010). Specifically, we derived the following competing hypotheses:

1) Similarity hypothesis: People with high self-control will express greater liking for other people with high self-control than for people with low self-control. People with low self-control will express greater liking for other people with low self-control than for people with high self-control.

2) Complementarity hypothesis: People with high self-control will express lower liking for other people with high self-control than for people with low self-control. People with low self-control will express lower liking for other people with low self-control than for people with high self-control.

3) Compensation hypothesis: People with low self-control will express greater liking for people with high self-control than for other people with low self-control. However, people with high self-control will express equally high levels of liking for people with high or low self-control.

4) Positivity hypothesis of self-control: Everyone will express greater liking for people with high self-control than for people with low self-control.

Moreover, liking has been found to depend on how well a person fits a stereotype: Probably the most salient category regarding stereotypes is the target person's sex (e.g., Hertel, Schütz, DePaulo, Morris, & Stucke, 2007). For example, anger is perceived more strongly in men than in women, whereas the opposite pattern has been found for sadness (Plant, Kling, & Smith, 2004). As men are stereotypically viewed as more aggressive than women and aggression is linked to losing one's temper, high aggressiveness may correspond with low self-control (Tangney et al., 2004). In line with this argument, men were found to display more explosive than defusing behavior (Campbell & Muncer, 2008). According to the German Federal Statistics

Office, the perpetrators of accidents that involve injuries to people caused by people who are under the influence of alcohol or drugs are comprised of 87% men and 13% women. The same has been found for speeding (78% men; Statistisches Bundesamt, 2019). Thus, compared with women, men should be perceived as having less self-control. We hypothesized that conforming to this stereotype would boost liking. Accordingly, we propose the following sex-stereotype hypothesis.

➢ Sex-stereotype hypothesis: Women with high self-control will be liked more than men with high self-control.

Halo Effects of High Self-Control

How much people like a person or how much they would like to be with him or her in a specific situation might not only depend on the person's apparent self-control. Halo effects (e.g., Nisbett & Wilson, 1977) suggest that high trait self-control, a trait that is generally evaluated in a positive manner, could lead people to assume that the person possesses other desirable traits, which in turn may influence how much they like the person. We used the continuum model (Fiske & Neuberg, 1990) to explore which traits may be seen as related to a person's trait self-control. The model is broad as it consists of eight scales comprised of adjectives that describe people and that can be arranged in terms of how much control a person exerts over others (i.e., assured-dominant vs. unassured-submissive) and how strongly a person strives for affiliation (i.e., cold-hearted vs. warm-agreeable).

As yet, only a few social perception studies investigated burdens of high self-control. For example, high self-control people are burdened by others' reliance (Koval et al., 2015). Fairness in supervisors (Cox, 2000; as cited by Tangney et al., 2004) and trust (Righetti & Finkenauer, 2011) are affected positively by high self-control. There was no evidence of negative effects of

self-control in either of the studies; instead, high self-control was *positively* associated with

perceived fairness (as found in the study by Cox) and trust (as found in both studies). There have

been similar findings involving school relationships: High-achieving students who probably have

high self-control were also regarded as being highly conscientious (Rentzsch, Schröder-Abé, &

Schütz, 2012). In view of the potential importance of halo effects and other imputed traits, we

added an exploratory measure of what other traits people with high or low self-control were

assumed to have.

Method

We investigated how hypothetical individuals with high versus low self-control are

perceived by others with respect to (a) how much people would like to be with these individuals

in different situations and (b) how much people like these individuals, and (c) how people rate

these individuals on a list of traits. We used vignettes in which several aspects were

systematically manipulated. As suggested by Simmons, Nelson, and Simonsohn (2012), we

report how we determined our sample size, all data exclusions (if any), all manipulations, and all

measures used in the study.

Participants

To achieve the planned sample size, we shared the online study's link via e-mail

messages and social networks. We incentivized people with the chance to win a 50€ Amazon

coupon. Participants who studied at the university at which the study was conducted could

alternatively receive course credit. For participants to be eligible for data analysis, they had to be

at least 18 years old and have completed the full study.

After we excluded participants who did not complete the questionnaire and participants

whose answers showed patterns such as simply checking the same number across all items, the

remaining sample size was $N = 285$ (71.9% women, mean age = 28.99 years). This far exceeded our planned sample size of approximately 200 participants. Thus, we conducted a sensitivity analysis for the term for which the study design was the least powerful, that is, within-between interactions (e.g., sex-stereotype hypothesis). Using a repeated-measures ANOVA with two groups and two measurements, the minimum effect size needed to be $f = 0.107$ ($\alpha = \beta = .05$, $N = 285$, correlation among repeated measures = .5, $\varepsilon = 1$), which is close to a small ($f = 0.100$) effect. The sensitivity of the analysis given our sample size was thus good. The power analyses were conducted with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Materials

Participants were presented with a vignette, which was based loosely on past work (Hertel et al., 2007). The vignette contained the person's name, age, sex, place of residence, hobbies, and a description of the person's level of self-control. We varied the vignette's characteristics in a 2 (self-control: high vs. low) × 2 (sex: male vs. female) × 2 (method factor: items used to describe the person) × 2 (order: high-low vs. low-high) design. The first factor was administered within-subjects.

The manipulation of the stimulus person's self-control used items from the German version of the Brief Self-Control scale (Bertrams & Dickhäuser, 2009; Tangney et al., 2004). We adopted this procedure from Shea et al. (2013). We excluded one item due to a low item-total correlation ($r < .3$; see also Bertrams & Dickhäuser, 2009), another because it referred to the respondent's self-perception, and a third because it invoked the respondent's perception of others. We then divided the remaining 10 items into two groups. To do so, we ordered them by their item-total correlations and assigned them to two groups using an ABBA scheme. The A-group items were thus very similar to the B-group items with respect to their mean item-total

correlations (r_a = .428, r_b = .414). Finally, to indicate whether the vignettes had high or low self-control, we negated the items. For example, "Sometimes I can't stop myself from doing something, even if I know it is wrong" was modified to be "Sarah can usually stop herself from doing something she knows is wrong" (high self-control) or "Sarah usually can't stop herself from doing something even when she knows it is wrong."

The person was either called Jan (male) or Sarah (female), which are the most common names for German people born in 1992. Their age was 25 in both versions (the study was conducted in early 2018). In order to reduce demand characteristics, we also varied their places of residence, occupation, and hobbies, for example, Jan lived in Duisburg, Germany, was an IT specialist, and liked playing football and guitar, and Sarah lived in Bochum, Germany, was a laboratory assistant, and liked aerobics and playing the keyboard. Thereby, we chose similar options: For example, the place of residence was chosen to have a similar number of inhabitants, the occupation was chosen to have a similar level of income, and the hobbies were chosen to be similar with respect to one of them referring to playing an instrument and the other a team sport.

To assess whether the manipulation of self-control was successful, participants were asked to indicate the presented person's level of self-control after each vignette (manipulation check). As dependent variables, we measured how likeable the people in the vignettes were. We asked the participants how much they would like to be in certain situations with the people described in the vignettes. Situations were either duty situations or socializing situations based on the taxonomy by Rauthmann et al. (2014). Which 5×2 situations were presented in which category was determined by a pretest for which we developed 7×2 situations. The duty situations featured (a) *being in a study group or team*, (b) *proofreading an application*, (c) *transcribing a presentation*, (d) *asking about a presentation*, (e) *doing the final cleaning of an apartment*, (f)

getting advice for a decision, and (g) *telling about a personal problem*. The socializing situations were (a) *spending the holidays*, (b) *having a conversation at a party*, (c) *partying*, (d) *going for a walk and having a nice conversation*, (e) *eating away from home*, (f) *spending an evening playing games,* (g) *cooking*. After pretesting, Situations 1 to 5 were selected for the main study, respectively. The socializing situations thus all involved having fun and pleasant, short-term interactions, as opposed to having a long-term relationship in which one might have to depend on the other person.

To assess possible halo effects, participants completed the adjective scales by Jacobs and Scholl (2005), which are based on the interpersonal theory of psychiatry by Sullivan (1953/2013). Information about the adequacy of the reliability and validity for the Interpersonal Adjective List can be found in Jacobs and Scholl (2005, pp. 152-153). There are a total of eight categories (in pairs of opposites) with eight adjectives each. The categories were assured-dominant versus unassured-submissive, gregarious-extraverted versus aloof-introverted, warm-agreeable versus cold-hearted, and unassuming-ingenuous versus arrogant-calculating. Finally, participants completed the German translation of the Brief Self-Control Scale (Bertrams & Dickhäuser, 2009; Tangney et al., 2004). Information about the adequacy of the reliability and validity (e.g., strong correlation with students' grade point average) for the Brief Self-Control Scale can be found in Tangney et al. (2004, p. 287). For all scales, we used equidistant anchors ranging from 1 (*disagree*) to 5 (*agree*) as proposed by Rohrmann (1978).

Procedure

First, participants provided information about their age, sex, and educational status. Then they were presented with the vignette, the manipulation check, the interpersonal adjectives list, and the situations. They were asked how much they would like to be in each of the situations

with the person described in the vignette. The procedure was the same for both vignettes. At the end of the study, participants completed the Brief Self-Control Scale (Tangney et al., 2004), described what they thought the purpose of the study was, and provided further demographic data to check for similarity with the descriptions in the vignette (place of residence, occupation, hobbies, field of study, semester).

Results

The data were analyzed with SPSS (SPSS Inc., 2017). The analysis script and data can be found online (osf.io/j3fsy). A summary of the hypotheses and findings is provided in Table 2.

Data Quality Checks

Manipulation check. For the item "self-controlled," which we had inserted into the interpersonal adjectives list, the high self-control vignette ($M = 4.45$, $SD = 0.84$) received substantially higher ratings than the low self-control vignette ($M = 1.59$, $SD = 0.79$), $t(284) = 40.16$, $p < .001$, $dz = 2.379$. The manipulation was thus quite powerful and successful at creating impressions of the stimulus individuals as high or low in self-control.

We further tested whether the different items we used to describe the high and low self-control vignettes and the order of the two vignettes (high-low vs. low-high self-control) had an effect on the ratings. As ratings, we used the duty situation liking because it was most affected by the other variables and thus seemed to be the most sensitive. There were no order or method effects: Whether participants rated the high self-control vignette or the low self-control vignette first did not have an effect (i.e., no order effects), $F(1, 283) = 1.47$, $p = .226$, $\eta^2_p = .005$. The specific items, that is, the items from the Brief Self-Control Scale that were used to describe self-control, did not have an effect either, $F(1, 283) = 0.02$, $p = .892$, $\eta^2_p < .001$. There was also no

effect of the interaction between the level of self-control in the vignette and the items, $F(1, 283)$ = 0.61, p = .434, η^2_p = .002.

We checked the internal consistencies of all scales to make sure participants had completed the survey conscientiously. Cronbach's alpha for the Brief Self-Control Scale (13 items, nine reversed ones) was good, α = .842, and very close to the values reported by Tangney et al. (2004, p. 287; α = .83 and .85). Internal consistencies for the duty situations scale were excellent (five items, α = .930 for the Sarah vignette and α = .938 for the Jan vignette). The internal consistency of the Socializing Scale was slightly lower but still good (five items), α = .889 and α = .870, for the two vignettes, respectively. The interpersonal adjective scales all had acceptable or good internal consistencies with the lowest value for unassuming-ingenuous (Sarah vignette), α = .754.

Hypothesis Tests

Situation hypothesis. To test the situation hypothesis, which, in its strongest formulation claims that people prefer a partner with high self-control in duty situations and a partner with low self-control in socializing situations, we computed a 2 (situation type: duty vs. socializing) × (vignette self-control: high vs. low) repeated-measures ANOVA. There was no main effect of the type of situation, $F(1, 284)$ = 0.02, p = .897, η^2_p < .001. Overall, partners with high self-control were preferred regardless of the type of situation, $F(1, 284)$ = 198.16, p < .001, η^2_p = .411. However, there was a significant interaction between type of situation and vignette self-control: Consistent with the strong formulation of the situation hypothesis, people with high self-control were preferred as partners in duty situations but not in socializing situations, $F(1, 284)$ = 757.03, p < .001, η^2_p = .727. The self-control (dis)advantage was dz = 1.706 for duty situations and dz = -.394 for socializing situations. Thus, there was a significant preference for the person with high

self-control in the duty situation but also a significant preference for the person with low self-control in the socializing situation (both pairwise comparisons $p < .001$). The results are depicted in *Figure 1*.
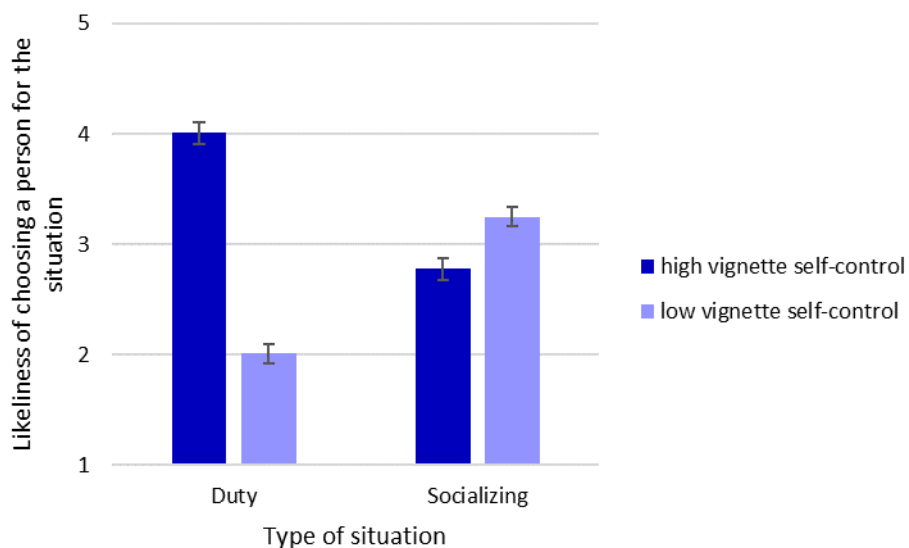

Figure 1. **Test of the situation hypothesis. Error bars represent 95% confidence intervals.**

Besides determining the influence of the type of situation and the vignette self-control, we included participants' self-control in the model for exploratory purposes. The results of the situation hypothesis test did not change. The only additional significant effect was a vignette self-control (duty vs. socializing) × participant self-control interaction effect, $F(1, 283) = 26.00$, $p < .001$, $\eta^2_p = .084$, such that participant self-control was positively correlated with the mean of all situation ratings for the high self-control vignettes, $r = .155$, and negatively correlated for the low self-control vignettes, $r = -.223$ (see *Figure 2*). To sum up, both vignette self-control *and* participant self-control interacted with the type of situation.
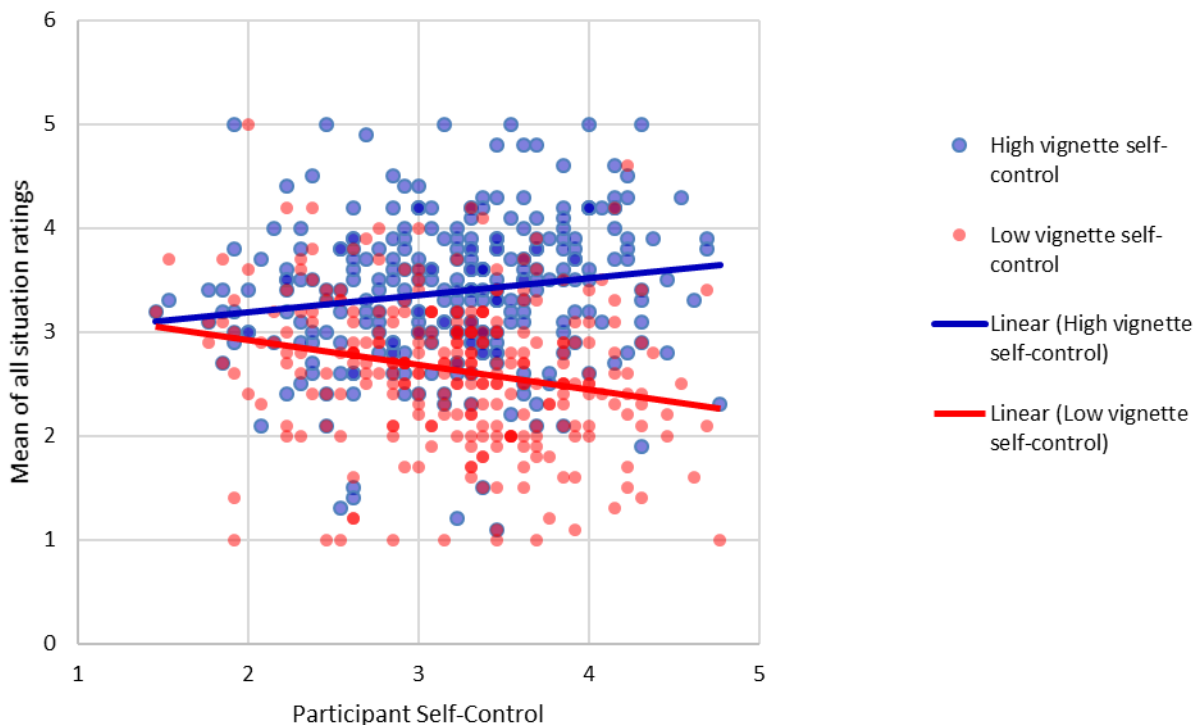
***Figure 2*. Interaction of participant self-control and vignette self-control on how likely participants were to choose the person described in the vignette as a partner in a variety of situations. The solid lines represent regression lines for the respective conditions of vignette self-control (blue: high; red: low).**

Liking hypothesis. We tested whether people with high self-control are liked more than people with low self-control and how this interacted with participants' own self-control using an ANOVA with vignette self-control (high vs. low) as a within-subjects factor and participant self-control as a metric factor. There was a main effect of vignette self-control, $F(1, 283) = 10.43$, $p = .001$, $\eta^2_p = .036$, such that people with high self-control were liked less than people with low self-control, a finding that was visible only when participants' self-control was controlled for ($M_{\text{liking vignette high self-control}} = 3.34$, $SD = 0.13$; $M_{\text{liking vignette low self-control}} = 3.38$, $SD = 0.12$, $dz = 0.155$) and not significant when no other variables were controlled for, $t(284) = 0.49$, $p = .625$, $dz = 0.041$. Participants' self-control had no main effect on the ratings, $F(1, 283) = 0.03$, $p = .863$, $\eta^2_p < .001$.

In line with the similarity hypothesis, participants' self-control moderated the effect of vignette self-control, $F(1, 283) = 10.20$, $p = .002$, $\eta^2_p = .035$. The results are depicted in *Figure 3*. People expressed the highest liking for people most similar to themselves on self-control.
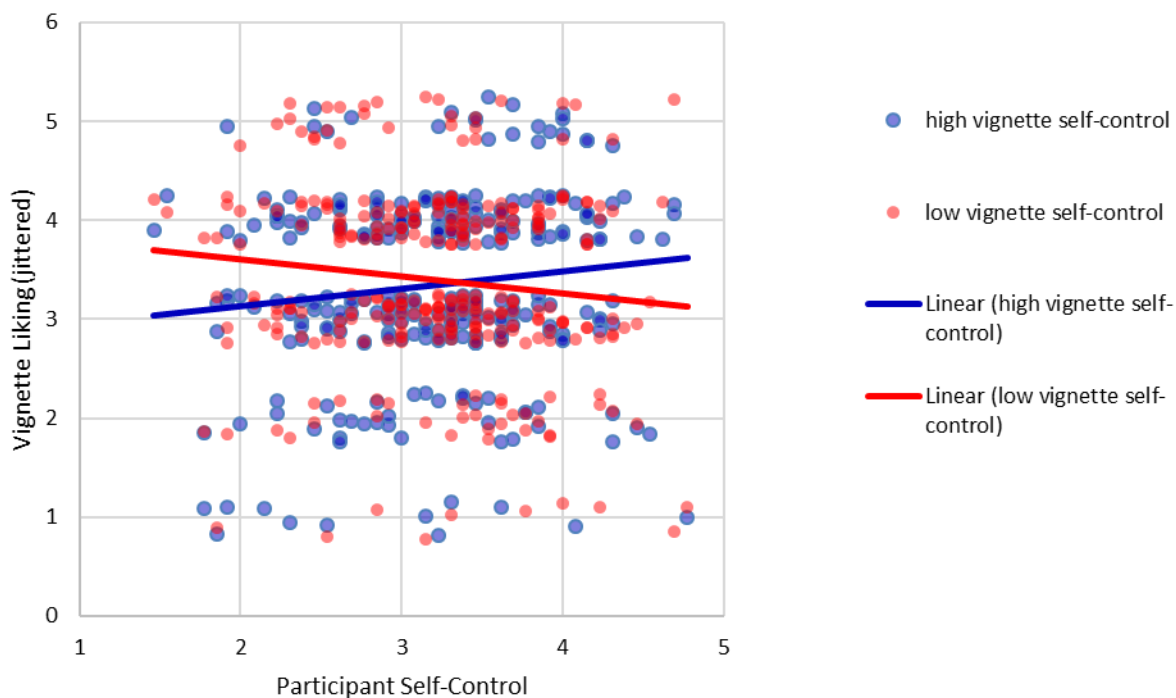

Figure 3. **Test of the liking hypothesis. Liking values are jittered for readability (i.e., random noise has been added to the scores to prevent overlap and enhance readability). The solid lines represent regression lines for the respective conditions of vignette self-control (blue: high; red: low).**

Sex-stereotype hypothesis. We tested whether a stereotype match (e.g., a man with low self-control or a woman with high self-control) had a positive effect on the extent to which participants liked that person. We computed a 2 (vignette sex: male vs. female) × 2 (stereotype match: yes vs. no) ANOVA. There was no main effect of vignette self-control when we did not control for respondents' self-control, $F(1, 283) = 0.24$, $p = .622$, $\eta^2_p < .001$. There was also no main effect of stereotype match, $F(1, 283) = 0.61$, $p = .435$, $\eta^2_p = .002$. Finally, there was no

interaction between vignette self-control and whether the vignettes matched the stereotypes, $F(1, 283) = 0.10$, $p = .750$, $\eta^2_p < .001$. The results did not support the sex-stereotype hypothesis (see also *Figure 4*).
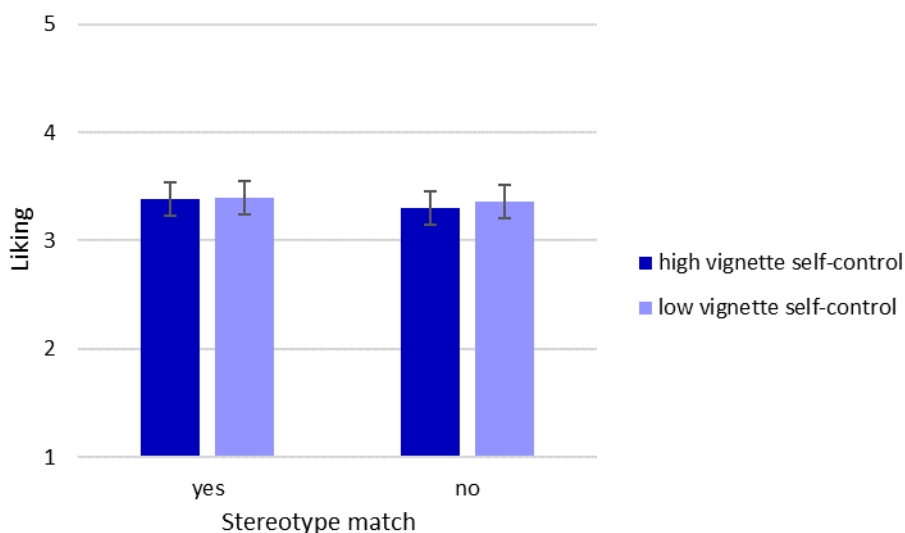

Figure 4. **Sex-stereotype hypothesis test. Error bars represent 95% confidence intervals.**

Exploratory Analyses

As we had not proposed specific hypotheses regarding halo effects of self-control on other traits, we investigated this aspect in an exploratory fashion. We ran eight *t* tests (one for each dimension) to determine the effect of vignette self-control. We controlled for an inflation of the alpha level by adjusting the significance level to $\alpha = .05/8 = .00625$ (Bonferroni correction). Results are depicted in *Figure 5* and Table 1. To summarize, almost all facets were sensitive to the manipulations of vignette self-control. People with high self-control were perceived as

dominant, unassuming, and introverted, whereas people with low self-control were perceived as

arrogant, cold-hearted, and extraverted[11].

Table 1

Differences in Vignettes (High vs. Low Self-Control) on the Dimensions of the Interpersonal
Adjective List (IAL)

Dimension	$M_{high\ vs.\ low}$	SD	dz	p
Assured-Dominant (PA)	0.66	1.06	0.619	<.001
Arrogant-Calculating (BC)	-0.20	0.96	-0.206	.001
Cold-Hearted (DE)	-0.24	0.80	-0.298	<.000
Aloof-Introverted (FG)	0.51	1.11	0.464	<.001
Unassured-Submissive (HI)	-0.07	1.16	-0.058	.325
Unassuming-Ingenuous (JK)	0.36	1.00	0.360	<.001
Warm-Agreeable (LM)	0.12	1.01	0.118	.047
Gregarious-Extraverted (NO)	-0.20	0.99	-0.201	<.001

---

[11] This is surprising given that dominance and introversion are negatively correlated

according to the circumplex model. We tested the circumplex model's proposed negative

relationships between the respective factors PA-HI, BC-JK, DE-LM, and FG-NO and found the

correlations to be much smaller than those reported by Jacobs and Scholl (2005, p. 146). The

strongest correlations were between assured-dominant (PA) and unassured-submissive (HI),

$r(283) = -.483$, $p < .001$ for the Jan vignettes and $r(283) = -.514$, $p < .001$ for the Sarah vignettes

(Jacobs and Scholl , 2005, p. 146 reported $r[332] = -.84$ for self-reports). However, correlations

between arrogant-calculating (BC) and unassuming-ingenuous (JK) were very small, $r(283) = -$

.099, $p = .094$, for the Jan vignettes and $r(283) = -.008$, $p = .897$ for the Sarah vignettes (Jacobs

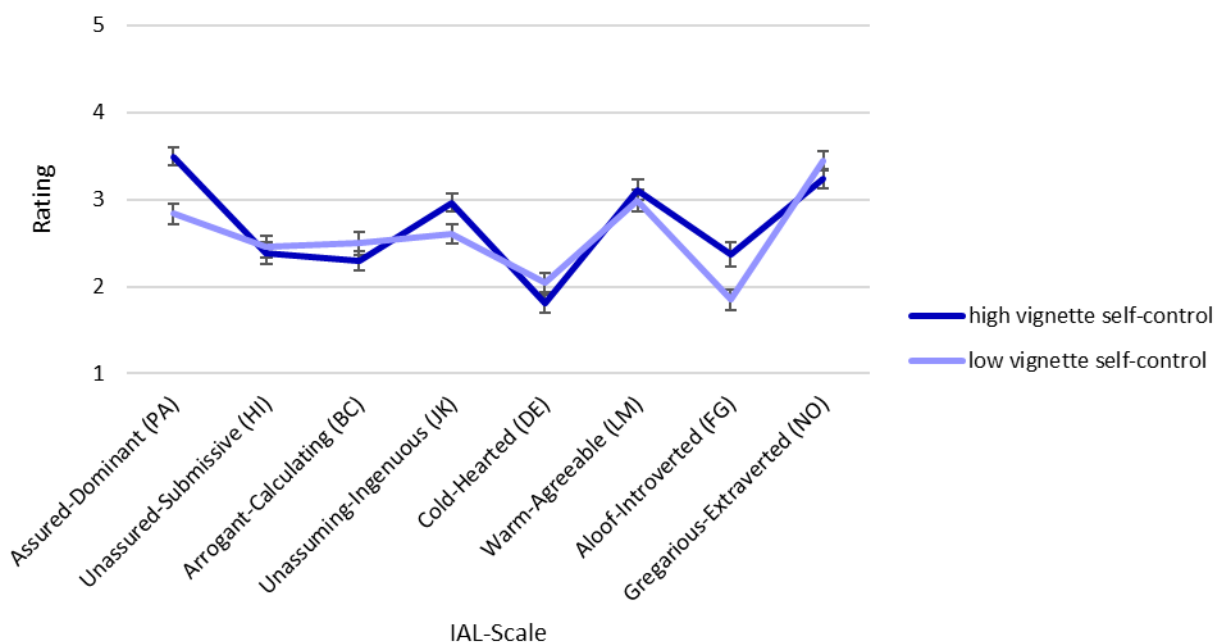and Scholl , 2005, p. 146 reported $r[332] = -.66$ for self-reports).

Figure 5. **Summary of IAL ratings for the high and low self-controlled hypothesis. Opposite dimensions (e.g., PA vs. HI) have been juxtaposed. Error bars represent 99.375% confidence intervals (Bonferroni corrected).**

Table 2

Summary of the Hypotheses and Results

Partner in a situation depends…	Liking of a person depends…	A sex stereotype – self-control fit…
H1: Strongly on the situation	*H1: On the similarity between respondent and target*	H1: Boosts liking
H2: Weakly on the situation	H2: On complementarity between the respondent and target	
H3: Positivity: on the partner's self-control	H3: Compensation: On the total self-control of respondent and target	*Has no effect*

On the similarity between respondent and partner	H4: Positivity: on the target's self-control being high
	Negativity: On the target's self-control being low

Note. Hypotheses that are supported by theories are marked with "H" and the respective number. Hypotheses that were supported by the data are written in italics.

Discussion

A broad and high-powered vignette study provided detailed evidence for a wide range of effects of a person's self-control on social perception. Contrary to the view that high self-control is always better than low self-control (positivity hypothesis of self-control), we found that the type of person that is preferred (a) depends on the type of situation one is confronted with, (b) depends on similarity, and (c) does not depend on gender stereotypes. In terms of duty situations (e.g., getting advice for a decision), people with high self-control were the preferred partners, whereas in socializing situations (e.g., going for a walk, attending a party, having a nice conversation), people with low self-control were the preferred partners. Also contradicting the positivity hypothesis of self-control, we found that how much people liked people with low or high self-control depended on their own self-control. More specifically, people with high self-control preferred other people with high self-control, and people with low self-control preferred other people with low self-control, a trend that is consistent with similarity being a major factor for determining attraction in nonromantic contexts.

There was an apparent contradiction of the main effect of self-control: When the task was to choose a partner for socializing and duty situations, participants preferred partners with high self-control overall (high vignette self-control). However, when we asked for how much participants liked the vignettes *and* participants' self-control was controlled for, low vignette self-control received higher liking scores overall. Note that choosing somebody in a situation differs from liking somebody and that the overall mean effect across all 10 situations depended on these situations and could easily be different for other situations. For example, the duty situations might have been more diagnostic than the sociality situations, but at the same time, they might occur less often and thus weigh less.

Although there is evidence for stereotypical male behavior being less self-controlled than female behavior, and there are theoretical grounds for predicting that a stereotype match would be positively associated with how much a person likes another person, we found no such effect: Whether men or women were presented as high or low in self-control had no differential effect on liking.

We found that the similarity effect and the situation effect occurred simultaneously and independently and dominated the main effect of vignette self-control (positivity hypothesis) by far. That is, the choice of a partner depends on the situation that the partner is needed for (situation hypothesis), the resemblance between the partner and the person who is choosing the partner (complementarity hypothesis), but not or only to a very small degree on the self-control of the partner being high (positivity hypothesis). Future research should consider these aspects of social perception in relation to self-control. We did find an advantage of people with high self-control over people with low self-control when people look for partners in duty situations. Despite this corroboration of the positivity hypothesis, other factors such as similarity and the

situation seem more important. This is surprising given that the definition of self-control includes aspects of social desirability. In other words, whether one likes somebody who is socially desirable (i.e., has high self-control) also depends on the perceiver's self-control. And whether one wants to work with somebody who is socially desirable depends on the situation. Ironically, in what we termed *socializing situations* (e.g., going on vacation together or partying), vignette characters with low self-control (i.e., a socially undesirable trait) were preferred over vignette characters with high self-control.

Finally, and more in line with the positivity hypothesis of self-control, an exploratory approach revealed that people with high self-control are associated with a wide range of other socially desirable traits, such as being assured, unassuming, and agreeable. Some of the results are seemingly contradictory: Although some factors theoretically correlate negatively (e.g., assured-dominant and gregarious-extraverted), differences between high and low self-controlled vignettes are not in the same direction with these factors. This could be due to patterns that are different in that population, possibly due to trans situational variability, than in a standard population. That is, people very high and low in self-control or the perceptions about these people could show other patterns of correlations between the dimensions that differ from those in a standard population.

Limitations

In our outline of the social perception of people with high self-control, we considered different kinds of situations, individual differences in the participants, sex stereotypes, and a wide range of traits. However, several factors merit further investigation.

First, we used vignettes only. Our setting was thus highly artificial, and the large effect sizes were probably due to the extreme manipulations. If actual people with less extreme differences were used, we would expect smaller effects.

Second, our study may have been affected by demand characteristics. It was evident that we were interested in perceptions that were based on the vignettes and on stereotypes. Even though people are not defenseless against demand characteristics (Hofmann, Gschwendner, Friese, Wiers, & Schmitt, 2008) judgments in everyday life may differ from findings in an experiment. The ratings of the vignettes might be subject to demand characteristics whereas spontaneous responses might not. Although tasks such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) have been criticized heavily (e.g., Schimmack, 2019), they allow for an assessment of such spontaneous reports (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) and can thus be helpful to avoid the problem of demand characteristics.

Third, our sample and vignettes targeted 20- to 30-year-old German people. Academic and social goals may vary across the lifespan and cultures, moderate the importance of duty and socializing situations, and thus affect their weight in choosing interaction partners and evaluating others. Even what is socially desirable might depend on these factors. For example, an interesting approach could be the longitudinal perspective on romantic relationships with respect to self-control. At the beginning of a romantic relationship, partners might exercise much more self-control in order to convince their partner of sustaining the rather fragile relationship than after a few months or years.

Although we identified situation type, similarity, and self-control as three parallel effects, with the last one being the weakest in our study, we cannot generalize their significance to contexts outside the laboratory. Generalization to other cultures where different traits are

considered socially desirable may not be possible either. Replications, especially in non-Western cultures, are needed. Due to the artificial setting and different manipulations and measures (i.e., manipulation of vignette self-control but no manipulation of respondent self-control), the effect sizes might be skewed.

Conclusion

Although self-control is usually understood as a socially desirable trait that is associated with functional behavior, we provided evidence against the simple assumption that more self-control is always better. Although there was a slight overall preference for people with high self-control, these preferences changed significantly (and even occasionally reversed) on the basis of similarity and type of situation. Our findings may be reassuring for people who have low self-control because they show that such people will sometimes be liked and chosen ahead of others. Sometimes, at least, it is advantageous to lack self-discipline.

References

Asendorpf, J. B., Penke, L., & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long-term outcomes of speed-dating in a community sample. *European Journal of Personality*, *25*(1), 16–30. https://doi.org/10.1002/per.768

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The Sticky Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues. *Journal of Behavioral Decision Making*, *87*(1), 23. https://doi.org/10.1002/bdm.2022

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science*, *16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität [Measurement of Dispositional Self-Control Capacity]. *Diagnostica*, *55*(1), 2–10. https://doi.org/10.1026/0012-1924.55.1.2

Böhm, R., Schütz, A., Rentzsch, K., Körner, A., & Funke, F. (2010). Are we looking for positivity or similarity in a partner's outlook on life? Similarity predicts perceptions of social attractiveness and relationship quality. *The Journal of Positive Psychology*, *5*(6), 431–438. https://doi.org/10.1080/17439760.2010.534105

Boman, J. H. (2016). Do Birds of a Feather Really Flock Together? Friendships, Self-Control Similarity and Deviant Behaviour. *British Journal of Criminology*, *38*, azw067. https://doi.org/10.1093/bjc/azw067

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality & Social Psychology*, *1*(6), 659–663. https://doi.org/10.1037/h0022073

BZgA, S. (2019). Verbreitung des regelmäßigen Alkoholkonsums in Deutschland nach

   Altersgruppen und Geschlecht im Jahr 2018 [Regular alcohol consumption in Germany by

   Age and Sex in 2018]. Retrieved from

   https://de.statista.com/statistik/daten/studie/443840/umfrage/verbreitung-des-regelmaessigen-

   alkoholkonsums-nach-altersgruppen-und-geschlecht/

Campbell, A., & Muncer, S. (2008). Intent to harm or injure? Gender and the expression of

   anger. *Aggressive Behavior*, *34*(3), 282–293. https://doi.org/10.1002/ab.20228

Carli, L. L., Ganley, R., & Pierce-Otay, A. (1991). Similarity and Satisfaction in Roommate

   Relationships. *Personality and Social Psychology Bulletin*, *17*(4), 419–426.

   https://doi.org/10.1177/0146167291174010

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social

   Psychology Review : An Official Journal of the Society for Personality and Social

   Psychology, Inc*, *13*(1), 62–72. https://doi.org/10.1177/1088868308329378

Dewall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depletion makes the

   heart grow less helpful: Helping as a function of self-regulatory energy and genetic

   relatedness. *Personality & Social Psychology Bulletin*, *34*(12), 1653–1662.

   https://doi.org/10.1177/0146167208323981

Dickman, S. J., & Meyer, D. E. (1988). Impulsivity and Speed-Accuracy Tradeoffs in

   Information Processing. *Journal of Personality & Social Psychology*. (54), 274–290.

Dryer, D. C., & Horowitz, L. M. (1997). When do opposites attract? Interpersonal

   complementarity versus similarity. *Journal of Personality and Social Psychology*, *72*(3), 592–

   603. https://doi.org/10.1037/0022-3514.72.3.592

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. *Behavior*

    Research Methods, *39*(2), 175–191. Retrieved from

    http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-

    Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower3-BRM-Paper.pdf

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-

    Based to Individuating Processes: Influences of Information and Motivation on Attention and

    Interpretation, *23*, 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual

    differences in implicit cognition: The implicit association test. *Journal of Personality and*

    Social Psychology, *74*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Hahn, C. M., & Campbell, L. J. (2016). Birds of a Feather Laugh Together: An Investigation of

    Humour Style Similarity in Married Couples. *Europe's Journal of Psychology*, *12*(3), 406–

    419. https://doi.org/10.5964/ejop.v12i3.1115

Hertel, J., Schütz, A., DePaulo, B. M., Morris, W. L., & Stucke, T. S. (2007). She's single, so

    what?: How are singles perceived compared with people who are married? *Zeitschrift Für*

    Familienforschung, *19*(2), 139–158. Retrieved from

    https://nbnresolving.org/urn:nbn:de:0168-ssoar-58080

Hobbes, T., & Brooke, C. (1651/2017). *Leviathan*. *Penguin classics*. [Harmondsworth,

    Meddlesex]: Penguin Books.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis

    on the correlation between the implicit association test and explicit self-report measures.

Personality & Social Psychology Bulletin, *31*(10), 1369–1385.

https://doi.org/10.1177/0146167205275613

Hofmann, W., Gschwendner, T., Friese, M., Wiers, R. W., & Schmitt, M. (2008). Working

memory capacity and self-regulatory behavior: Toward an individual differences perspective

on behavior determination by automatic versus controlled processes. *Journal of Personality*

and Social Psychology, *95*(4), 962–977. https://doi.org/10.1037/a0012705

Ináncsi, T., Láng, A., & Bereczkei, T. (2016). A Darker Shade of Love: Machiavellianism and

Positive Assortative Mating Based on Romantic Ideals. *Europe's Journal of Psychology*,

12(1), 137–152. https://doi.org/10.5964/ejop.v12i1.1007

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL) [Interpersonal Adjectivelist

(IAL)]. *Diagnostica*, *51*(3), 145–155. https://doi.org/10.1026/0012-1924.51.3.145

Kiesler, D. J. (1983). The 1982 Interpersonal Circle: A taxonomy for complementarity in human

transactions. *Psychological Review*, *90*(3), 185–214. https://doi.org/10.1037/0033-

295X.90.3.185

Koval, C. Z., vanDellen, M. R., Fitzsimons, G. M., & Ranby, K. W. (2015). The burden of

responsibility: Interpersonal costs of high self-control. *Journal of Personality and Social*

Psychology, *108*(5), 750–766. https://doi.org/10.1037/pspi0000015

Lösch, T., & Rentzsch, K. (2018). Linking Personality with Interpersonal Perception in the

Classroom: Distinct Associations with the Social and Academic Sides of Popularity. *Journal*

of Research in Personality, *75*, 83–93. https://doi.org/10.1016/j.jrp.2018.06.003

Maszk, P., Eisenberg, N. G., & Guthrie, I. K. (1999). Relations of Children's Social Status to

Their Emotionality and Regulation: A Short-Term Longitudinal Study. *Merrill-Palmer*

Quarterly, *45*, 468–492. Retrieved from https://www.jstor.org/stable/23092582

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922. https://doi.org/10.1177/0265407508096700

Morell, M. A., Twillman, R. K., & Sullaway, M. E. (1989). Would a Type A Date Another Type A?: Influence of Behavior Type and Personal Attributes In the Selection of Dating Partners1. *Journal of Applied Social Psychology*, *19*(11), 918–931. https://doi.org/10.1111/j.1559-1816.1989.tb01229.x

Morry, M. M. (2005). Relationship satisfaction as a predictor of similarity ratings: A test of the attraction-similarity hypothesis. *Journal of Social and Personal Relationships*, *22*(4), 561–584. https://doi.org/10.1177/0265407505054524

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037//0022-3514.35.4.250

Pelkner, A.-K., & Boehnke, K. (2003). Streber als Leistungsverweigerer? [Do Nerds Refuse to Achieve?]. *Zeitschrift Für Erziehungswissenschaft*, *6*(1), 106–125. https://doi.org/10.1007/s11618-003-0007-4

Plant, E. A., Kling, K. C., & Smith, G. L. (2004). The Influence of Gender and Social Role on the Interpretation of Facial Expressions. *Sex Roles*, *51*(3/4), 187–196. https://doi.org/10.1023/B:SERS.0000037762.10349.13

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., . . . Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy

of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rentzsch, K., Schröder-Abé, M., & Schütz, A. (2012). Being Called a 'Streber': The Roles of Personality and Competition in the Labelling of Academically Oriented Students. *European Journal of Personality*, *71*(4), n/a-n/a. https://doi.org/10.1002/per.1884

Rentzsch, K., Schütz, A., & Schröder-Abé, M. (2011). Being Labeled Nerd : Factors That Influence the Social Acceptance of High-Achieving Students. *The Journal of Experimental Education*, *79*(2), 143–168. https://doi.org/10.1080/00220970903292900

Ridder, D. T. D. de, Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *16*(1), 76–99. https://doi.org/10.1177/1088868311418749

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology*, *100*(5), 874–886. https://doi.org/10.1037/a0021827

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozalwissenschaftliche Forschung [Empirical Studies on the Development of Response Scales for the Social Sciences]. *Zeitschrift Für Sozialpsychologie*, *9*(3), 222–245. Retrieved from http://www.f3.htw-berlin.de/Professoren/Pruemper/pdf/Antwortskalen.pdf

Schimmack, U. (2019). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, 1745691619863798. https://doi.org/10.1177/1745691619863798

Schröder-Abé, M., & Schütz, A. (2011). Walking in each other's shoes: Perspective taking

mediates effects of emotional intelligence on relationship quality. *European Journal of*

Personality, *25*(2), 155–169. https://doi.org/10.1002/per.818

Shea, C. T., Davisson, E. K., & Fitzsimons, G. M. (2013). Riding other people's coattails:

Individuals with low self-control value self-control in other people. *Psychological Science*,

24(6), 1031–1036. https://doi.org/10.1177/0956797612464890

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic*

Journal. Advance online publication. https://doi.org/10.2139/ssrn.2160588

IBM SPSS Statistics for Windows, Version 25.0 (Version 25.0) [Computer software] (2017).

Armonk, NY: IBM Corp.

Statistisches Bundesamt (2019). Unfälle unter dem Einfluss berauschender Mittel im

Straßenverkehr [Traffic Accidents under the Influence of Narcotics]. Retrieved from

https://www.destatis.de/DE/Themen/Gesellschaft-

Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/unfaelle-alkohol-

5462404187004.pdf

Sullivan, H. S. (1953/2013). *The interpersonal theory of psychiatry*: Routledge.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good

Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of*

Personality, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and

submissive nonverbal behavior. *Journal of Personality & Social Psychology*, *84*(3), 558–568.

https://doi.org/10.1037//0022-3514.84.3.558

Vohs, K. D., Finkenauer, C., & Baumeister, R. F. (2011). The Sum of Friends' and Lovers' Self-

Control Scores Predicts Relationship Quality. *Social Psychological and Personality Science*,

2(2), 138–145. https://doi.org/10.1177/1948550610385710

Biographical Statement

LR is a PhD student at the University of Bamberg and the Harz University of Applied Sciences. His research focuses on assimilation and contrast. He also studies self-control and cognitive biases.

JE studied psychology at the University of Bamberg. This project is rooted in her Bachelor's thesis.

AS is a personality psychologist based at the University of Bamberg. Her research focuses on personality in social relationships. She studies self-esteem, narcissism and emotional intelligence in social interactions and devices tools for training and assessment.

RFB is a social psychologist based at the University of Queensland. He explores how we think about the self, and why we feel and act the way we do. He is especially known for his work on the subjects of willpower, self-control, and self-esteem, and how they relate to human morality and success.

Funding/Financial Support

Acknowledgements

References

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R [Benjamin R.], . . . Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*(1), 4–6. https://doi.org/10.1038/s41562-019-0772-6

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1).

Ames, D. R., & Mason, M. F. (2015). Tandem anchoring: Informational and politeness effects of range offers in social exchange. *Journal of Personality and Social Psychology*, *108*(2), 254–274. https://doi.org/10.1037/pspi0000016

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

Bahník, Š., Englich, B., & Strack, F. (2017). Anchoring effect. In R. F. Pohl (Ed.), *A Psychology Press book. Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 223–241). Routledge.

Bahník, Š., Houdek, P., Vrbova, L., & Hájek, J. (2018). Variations on anchoring. Advance online publication. https://doi.org/10.31234/osf.io/skmqg

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2017). The Sticky Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues. *Journal of Behavioral Decision Making*, *87*(1), 23. https://doi.org/10.1002/bdm.2022

Bastian, C. C. von, Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P.,
Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M.,
Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T.,
Gustavson, D. E., Koch, I., . . . Wiemers, E. A. (2020). Advancing the understanding of
individual differences in attentional control: Theoretical, methodological, and analytical
considerations. Advance online publication. https://doi.org/10.31234/osf.io/x3b9k

Baumeister, R. F., & Vohs, K. D. (2016). Misguided Effort With Elusive Implications.
*Perspectives on Psychological Science : A Journal of the Association for Psychological
Science*, *11*(4), 574–575. https://doi.org/10.1177/1745691616652878

Bem, D., Duggan, M., Rabeyron, T., & Tressoldi, P. (2015). Feeling the future: A meta-analysis
of 90 experiments on the anomalous anticipation of random future events. *F1000Research.*
Advance online publication. https://doi.org/10.12688/f1000research.7177.1

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

Bishop, D. V. M., & Thompson, P. A. (2016). Problems in using p-curve analysis and text-
mining to detect rate of p-hacking and evidential value. *PeerJ*, *4*, e1715.
https://doi.org/10.7717/peerj.1715

Boyle, G. J., Matthews, G., & Saklofske, D. H. (2008). *The SAGE handbook of personality
theory and assessment* (1st ed.). SAGE Publications.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R.,
Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The Replication Recipe:
What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*,
217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing Research for Application. *Journal of Consumer Research*, *8*(2), 197. https://doi.org/10.1086/208856

Caputo, A. (2014). Relevant information, personality traits and anchoring effect. *International Journal of Management and Decision Making*, *13*(1), Article 58470, 62. https://doi.org/10.1504/IJMDM.2014.058470

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*(4), 223–242. https://doi.org/10.1002/bdm.3960070402

Cheek, N. N., & Norem, J. K. (2018). On Moderator Detection in Anchoring Research: Implications of Ignoring Estimate Direction. *Collabra: Psychology*, *4*(1), 12. https://doi.org/10.1525/collabra.125

Cheek, N. N., & Norem, J. K. (2019). Are Big Five Traits and Facets Associated With Anchoring Susceptibility? *Social Psychological and Personality Science*, *92*, 194855061983700. https://doi.org/10.1177/1948550619837001

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Taylor and Francis.

Cowan, N., Belletier, C., Doherty, J. M., Jaroslawska, A. J., Rhodes, S., Forsberg, A., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2020). How Do Scientific Views Change? Notes From an Extended Adversarial Collaboration. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *15*(4), 1011–1025. https://doi.org/10.1177/1745691620906415

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, *21*(3), 241–251. https://doi.org/10.1002/bdm.586

Czerwonka, M. (2017). Anchoring and Overconfidence: The Influence of Culture and Cognitive Abilities. *International Journal of Management and Economics*, *53*(3), 48–66. https://doi.org/10.1515/ijme-2017-0018

Dogerlioglu-Demir, K., & Koçaş, C. (2015). Seemingly incidental anchoring: the effect of incidental environmental anchors on consumers' willingness to pay. *Marketing Letters*, *26*(4), 607–618. https://doi.org/10.1007/s11002-014-9295-0

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS One*, *7*(1), e29081. https://doi.org/10.1371/journal.pone.0029081

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychological Science*, *31*(7), 792–806. https://doi.org/10.1177/0956797620916786

Englich, B., & Soder, K. (2009). Moody experts - How mood and expertise influence judgmental anchoring. *Judgment and Decision Making*, *4*, 41–50.

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

Epley, N., & Gilovich, T. (2010). Anchoring unbound. *Journal of Consumer Psychology*, *20*(1), 20–24. https://doi.org/10.1016/j.jcps.2009.12.005

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133. https://doi.org/10.1016/j.ijforecast.2009.02.005

Fechner, G. T. (1876/2013). *Vorschule der Aesthetik*. Cambridge University Press. https://doi.org/10.1017/CBO9781139854597

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a Within-Subject, Repeated-Measures Ego-Depletion Paradigm. *Social Psychology*, *49*(5), 271–286. https://doi.org/10.1027/1864-9335/a000348

Frech, M.-L., Loschelder, D. D., & Friese, M. (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. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology. General*, *141*(1), 124–133. https://doi.org/10.1037/a0024006

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2018). Is Ego Depletion Real? An Analysis of Arguments. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc.* Advance online publication. https://doi.org/10.1177/1088868318762183

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Furnham, A., Boo, H. C., & McClelland, A. (2012). Individual Differences and the Susceptibility to the Influence of Anchoring Cues. *Journal of Individual Differences*, *33*(2), 89–93. https://doi.org/10.1027/1614-0001/a000076

Galton, F. (1884). Measurement of character. *Fortnightly*, *36*(220), 179–185.

Garrison, K. E., Finley, A. J., & Schmeichel, B. J. (2019). Ego Depletion Reduces Attention Control: Evidence From Two High-Powered Preregistered Experiments. *Personality & Social Psychology Bulletin*, *45*(5), 728–739. https://doi.org/10.1177/0146167218796473

Gervais, W. M., McKee, S. E., & Malik, S. (2020). Do Religious Primes Increase Risk Taking? Evidence Against "Anticipating Divine Protection" in Two Preregistered Direct Replications

of Kupor, Laurin, and Levav (2015). *Psychological Science*, *31*(7), 858–864. https://doi.org/10.1177/0956797620922477

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*(4), 601–613. https://doi.org/10.1037/0022-3514.59.4.601

Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment and Decision Making*, *6*(8), 711–721.

Gordon, M., Viganola, D., Bishop, M., Chen, Y., Dreber, A., Goldfedder, B., Holzmeister, F., Johannesson, M., Liu, Y., Twardy, C., Wang, J., & Pfeiffer, T [Thomas] (2020). Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. *Royal Society Open Science*, *7*(7), 200566. https://doi.org/10.1098/rsos.200566

Grice, H. P. (1982). *Speech acts* (5. ed.). *Syntax and semantics: Vol. 3*. Academic Press.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R [Angela R.], Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., Ridder, D. T. D. de, Dewitte, S., . . . Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495–525. https://doi.org/10.1037/a0019486

Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., Urlichich, D., & Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology. General.* Advance online publication. https://doi.org/10.1037/xge0000644

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science*, *16*(7), 566–571. https://doi.org/10.1111/j.0956-7976.2005.01576.x

Helson, H. (1964). *Adaptation-level theory: an experimental and systematic approach to behavior.*

Hoenig, J. M., & Heisey, D. M. (2001). The Abuse of Power. *The American Statistician*, *55*(1), 19–24. https://doi.org/10.1198/000313001300339897

Hovland, C. I., Harvey, O. J., & Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *The Journal of Abnormal and Social Psychology*, *55*(2), 244–252. https://doi.org/10.1037/h0048480

Hoyningen-Huene, P. (2013). *Systematicity: The nature of science. Oxford studies in philosophy of science*. Oxford Univ. Press.

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

Inzlicht, M., & Friese, M. (2019). The Past, Present, and Future of Ego Depletion. *Social Psychology*, *50*(5-6), 370–378. https://doi.org/10.1027/1864-9335/a000398

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL) [Interpersonal Adjectivelist (IAL)]. *Diagnostica*, *51*(3), 145–155. https://doi.org/10.1026/0012-1924.51.3.145

Jacowitz, K. E., & Kahnemann, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166. https://doi.org/10.1177/01461672952111004

Jäger, M., & Heydasch, T. (2016). Hagener Matrizen-Test (HMT) (PSYNDEX Tests Review). In *Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Ed.), PSYNDEX. Trier: ZPID.*

Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science*, *19*(2), 121–127. https://doi.org/10.1111/j.1467-9280.2008.02057.x

Jasper, F., & Ortner, T. M. (2014). The Tendency to Fall for Distracting Information While Making Judgments. *European Journal of Psychological Assessment*, *30*(3), 193–207. https://doi.org/10.1027/1015-5759/a000214

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kahneman, D., & Knetsch, J. L. (1993). Strong influences and shallow inferences: An analysis of some anchoring effects. *Unpublished Manuscript. University of California, Berkeley*.

King, D. L. (1988). Assimilation is due to one perceived whole and contrast is due to two perceived wholes. *New Ideas in Psychology*, *6*(3), 277–288. https://doi.org/10.1016/0732-118X(88)90039-6

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J [Michael J.], Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D.,

Galliani, E. M., . . . Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., Babalola, M. T., Bahník, Š., Berkics, M., Bernstein, M. J [Michael Jason], Berry, D. R., Bialobrzeska, O., Bocian, K., Brandt, M., Busching, R., Cai, H., Cambier, F., . . . Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. Advance online publication. https://doi.org/10.31234/osf.io/9654g

Lakens, D [Daniel] (2014). Professors are Not Elderly: Evaluating the Evidential Value of Two Social Priming Effects Through P-Curve Analyses. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2381936

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

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T [Thomas], Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M., Derks, K., Wagenmakers, E.-J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., . . . Uhlmann, E. L [Eric L.] (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451–479. https://doi.org/10.1037/bul0000220

LeBel, E. P. (2015). A New Replication Norm for Psychology. *Collabra*, *1*(1). https://doi.org/10.1525/collabra.23

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, *15*(4), 371–379. https://doi.org/10.1037/a0025172

Leder, J., Foster, J. D., & Schütz, A. (2020). Willing to do what others won't: Narcissism, risk-taking, and social approval. *Personality and Individual Differences*, *153*, 109672. https://doi.org/10.1016/j.paid.2019.109672

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the big five: Are they within the HEXACO space? *Journal of Personality*, *73*(5), 1437–1463. https://doi.org/10.1111/j.1467-6494.2005.00354.x

Lewis, J., Gaertig, C., & Simmons, J. P. (2019). Extremeness Aversion Is a Cause of Anchoring. *Psychological Science*, *30*(2), 159–173. https://doi.org/10.1177/0956797618799305

Linden, A. H., & Hönekopp, J. (in press). Heterogeneity and Progress accepted. *Perspectives on Psychological Science*. https://osf.io/qsbrn/

Littrell, S., Fugelsang, J., & Risko, E. F. (2020). Not so fast: Individual differences in impulsiveness are only a modest predictor of cognitive reflection. *Personality and Individual Differences*, *154*, 109678. https://doi.org/10.1016/j.paid.2019.109678

Loschelder, D. D., Friese, M., Schaerer, M., & Galinsky, A. D. (2016). The Too-Much-Precision Effect. *Psychological Science*, *27*(12), 1573–1587. https://doi.org/10.1177/0956797616666074

Loschelder, D. D., Stuppi, J., & Trötschel, R. (2014). "€14,875?!": Precision Boosts the

    Anchoring Potency of First Offers. *Social Psychological and Personality Science*, *5*(4), 491–

    499. https://doi.org/10.1177/1948550613499942

Maniadis, Z., Tufano, F., & List, J. A. (2014). One Swallow Doesn't Make a Summer: New

    Evidence on Anchoring Effects. *American Economic Review*, *104*(1), 277–290.

    https://doi.org/10.1257/aer.104.1.277

Mason, M. F., Lee, A. J., Wiley, E. A., & Ames, D. R. (2013). Precise offers are potent anchors:

    Conciliatory counteroffers and attributions of knowledge in negotiations. *Journal of*

    Experimental Social Psychology, *49*(4), 759–763. https://doi.org/10.1016/j.jesp.2013.02.012

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median

    splits, Type II errors, and false–positive consumer psychology: Don't fight the power. *Journal*

    of Consumer Psychology, *25*(4), 679–689. https://doi.org/10.1016/j.jcps.2015.05.006

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-

    experience influences responses to anchoring cues. *Judgment and Decision Making*, *2*(1), 48–

    53.

Meub, L., Proeger, T., & Bizer, K. (2013). Anchoring: A Valid Explanation for Biased Forecasts

    When Rational Predictions are Easily Accessible and Well Incentivized? *SSRN Electronic*

    Journal. Advance online publication. https://doi.org/10.2139/ssrn.2307619

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., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and
    underlying mechanisms. *Organizational Behavior and Human Decision Processes*, *98*(2),
    133–143. https://doi.org/10.1016/j.obhdp.2004.12.002

Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of
    judgemental anchoring. *European Review of Social Psychology*, *10*(1), 135–167.
    https://doi.org/10.1080/14792779943000044

Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the
    anchoring paradigm: A selective accessibility model. *Journal of Experimental Social
    Psychology*, *35*, 136–164. https://doi.org/10.1037/e413782005-562

Mussweiler, T., & Strack, F. (2000a). Numeric Judgments under Uncertainty: The Role of
    Knowledge in Anchoring. *Journal of Experimental Social Psychology*, *36*(5), 495–518.
    https://doi.org/10.1006/jesp.1999.1414

Mussweiler, T., & Strack, F. (2000b). The use of category and exemplar knowledge in the
    solution of anchoring tasks. *Journal of Personality and Social Psychology*, *78*(6), 1038–1052.
    https://doi.org/10.1037//0022-3514.78.6.1038

Mussweiler, T., & Strack, F. (2001a). Considering The Impossible: Explaining The Effects of
    Implausible Anchors. *Social Cognition*, *19*(2), 145–160.
    https://doi.org/10.1521/soco.19.2.145.20705

Mussweiler, T., & Strack, F. (2001b). The semantics of anchoring. *Organizational Behavior and
    Human Decision Processes*, *86*(2), 234–255. https://doi.org/10.1006/obhd.2001.2954

Mussweiler, T., Strack, F., & Pfeiffer, T [Tim] (2000). Overcoming the Inevitable Anchoring
    Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and
    Social Psychology Bulletin*, *26*(9), 1142–1150. https://doi.org/10.1177/01461672002611010

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *349*(Suppl 1), aac4716. https://doi.org/10.1038/s41562-018-0522-1

Nebioglu, M., Konuk, N., Akbaba, S., & Eroglu, Y. (2012). The Investigation of Validity and Reliability of the Turkish Version of the Brief Self-Control Scale. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology*, *22*(4), 340–351. https://doi.org/10.5455/bcp.20120911042732

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology*, *69*, 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., Alshaif, N., Andringa, R., Aveyard, M., Babincak, P., Balatekin, N., Baldwin, S. A., Banik, G., Baskin, E., Bell, R., Białobrzeska, O., Birt, A. R [Angie R.], Boot, W. R., Braithwaite, S. R., . . . Zrubka, M. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *13*(2), 268–294. https://doi.org/10.1177/1745691618755704

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), 147–152. https://doi.org/10.1016/j.jebo.2009.04.018

Olsson-Collentine, A., Wicherts, J. M., & van Assen, M. A. L. M. (2020). Heterogeneity in direct replications in psychology and its association with effect size. *Psychological Bulletin.* Advance online publication. https://doi.org/10.1037/bul0000294

Open Science Collaboration (2015). Psychology: Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Parsons, S., Kruijt, A.-W., & Fox, E. (2018). Psychological Science needs a standard practice of reporting the reliability of cognitive behavioural measurements. Advance online publication. https://doi.org/10.17605/OSF.IO/6KA9Z

Perry, R., & Sibley, C. G. (2013). Seize and freeze: Openness to Experience shapes judgments of societal threat. *Journal of Research in Personality*, *47*(6), 677–686. https://doi.org/10.1016/j.jrp.2013.06.006

Peter, J. P., Churchill, J. G. A., & Brown, T. J. (1993). Caution in the Use of Difference Scores in Consumer Research. *Journal of Consumer Research*, *19*(4), 655. https://doi.org/10.1086/209329

Petty, R. E., DeMarree, K. G., Briñol, P., Horcajo, J., & Strathman, A. J. (2008). Need for cognition can magnify or attenuate priming effects in social judgment. *Personality & Social Psychology Bulletin*, *34*(7), 900–912. https://doi.org/10.1177/0146167208316692

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). Sara: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory (Hove, England)*, *11*(4-5), 337–356. https://doi.org/10.1080/09658210244000487

Popper, K. R. (1959/2008). *The Logic of Scientific Discovery* (Repr. 2008). *Routledge classics*. Routledge Classics; Routledge.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory

(IRT). *Journal of Behavioral Decision Making*, *29*(5), 453–469.

https://doi.org/10.1002/bdm.1883

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

Rohrer, J. M., Tierney, W., Uhlmann, E. L [Eric Luis], DeBruine, L. M., Heyman, T.,

Jones, B. C., Schmukle, S. C., Silberzahn, R., Willén, R. M., Carlsson, R., Lucas, R. E.,

Strand, J. F., Vazire, S., Witt, J. K., Zentall, T. R., Chabris, C., & Yarkoni, T. (2018). Putting

the Self in Self-Correction: Findings from the Loss-of-Confidence Project. Advance online

publication. https://doi.org/10.31234/osf.io/exmb2

Röseler, L., Schütz, A., Baumeister, R. F., & Starker, U. (2020). Does ego depletion reduce

judgment adjustment for both internally and externally generated anchors? *Journal of*

Experimental Social Psychology, *87*, 103942. https://doi.org/10.1016/j.jesp.2019.103942

Röseler, L., Schütz, A., Blank, P. A., Dück, M., Fels, S., Kupfer, J., Scheelje, L., & Seida, C.

(2020). No Evidence of Subliminal Anchoring: Two Close, Highly Powered, Preregistered,

and Failed Replication Attempts. Manuscript submitted for publication.

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., Schütz, A., & Starker, U. (2019). Cognitive Ability does not and cannot Correlate

with Susceptibility to Anchoring Effects. Advance online publication.

https://doi.org/10.31234/osf.io/bnsx2

Sailors, J. J., & Heyman, J. E. (2019). Similarity, multiple estimations, and the anchoring effect. *The Journal of General Psychology*, *146*(2), 200–215. https://doi.org/10.1080/00221309.2018.1551775

Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt. *Diagnostica*, *50*(4), 202–218. https://doi.org/10.1026/0012-1924.50.4.202

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: an inclusion/ exclusion model of assimilation and contrast effects in social judgement. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (Vol. 1991). Lawrence Erlbaum Associates. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-68883

Schweickart, O., Tam, C., & Brown, N. R. (2020). When "bad" is good: How evaluative judgments eliminate the standard anchoring effect. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000209

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

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, *55*(2), 150–155. https://doi.org/10.1037/h0048784

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2160588

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after P-Hacking. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2205186

Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology. General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2013). Anchoring is Not a False-Positive: Maniadis, Tufano, and List's (2014) 'Failure-to-Replicate' is Actually Entirely Consistent with the Original. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2351926

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller

(2015). *Journal of Experimental Psychology. General*, *144*(6), 1146–1152. https://doi.org/10.1037/xge0000104

Sjöberg, A., Sjöberg, S., & Forssen, K. (2006). *Predicting Job Performance*. Assessio International.

Smeesters, D., Yzerbyt, V. Y., Corneille, O., & Warlop, L. (2009). When do primes prime? The moderating role of the self-concept in individuals' susceptibility to priming effects on social behavior. *Journal of Experimental Social Psychology*, *45*(1), 211–216. https://doi.org/10.1016/j.jesp.2008.09.002

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, *47*(5), 385–395. https://doi.org/10.1016/j.paid.2009.04.008

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34. http://www.jstor.com/stable/2282137

Strack, F., Bahník, Š., & Mussweiler, T. (2016). Anchoring: Accessibility as a cause of judgmental assimilation. *Current Opinion in Psychology*, *12*, 67–70. https://doi.org/10.1016/j.copsyc.2016.06.005

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*(3), 437–446. https://doi.org/10.1037//0022-3514.73.3.437

Street, C. N. H., & Kingstone, A. (2017). Aligning Spinoza with Descartes: An informed

Cartesian account of the truth bias. *British Journal of Psychology (London, England : 1953)*,

108(3), 453–466. https://doi.org/10.1111/bjop.12210

Street, C. N. H., & Richardson, D. C. (2015). Descartes Versus Spinoza: Truth, Uncertainty, and

Bias. *Social Cognition*, *33*(3), 227–239. https://doi.org/10.1521/soco.2015.33.2.2

Teovanović, P. (2019). Individual Differences in Anchoring Effect: Evidence for the Role of

Insufficient Adjustment. *Europe's Journal of Psychology*, *15*(1), 8–24.

https://doi.org/10.5964/ejop.v15i1.1691

Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases:

Evidence against one-factor theory of rationality. *Intelligence*, *50*, 75–86.

https://doi.org/10.1016/j.intell.2015.02.008

Thorsteinson, T. J. (2011). Initiating Salary Discussions With an Extreme Request: Anchoring

Effects on Initial Salary Offers1. *Journal of Applied Social Psychology*, *41*(7), 1774–1792.

https://doi.org/10.1111/j.1559-1816.2011.00779.x

Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of

anchoring effects. *Cognitive Psychology*, *90*, 1–47.

https://doi.org/10.1016/j.cogpsych.2016.07.003

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

van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A

discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.

https://doi.org/10.1016/j.jesp.2016.03.004

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W.B. G. (2001). Implications of Attitude Change Theories for Numerical Anchoring: Anchor Plausibility and the Limits of Anchor Effectiveness. *Journal of Experimental Social Psychology*, *37*(1), 62–69. https://doi.org/10.1006/jesp.2000.1431

Welsh, M. B., Delfabbro, P. H., Burns, N. R., & Begg, S. H. (2011). Individual differences in anchoring: Numerical ability, education and experience. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33).

Welsh, M. B., Delfabbro, P. H., Burns, N. R., & Begg, S. H. (2014). Individual differences in anchoring: Traits and experience: How much more than Numerical Ability? *Learning and Individual Differences*, *29*, 131–140. https://doi.org/10.1016/j.lindif.2013.01.002

Wilson, T. D., Housten, C. E., & Brekke, N. (1996). A New Look at Anchoring Effects: Basic Anchoring and Its Antecedents. *Journal of Experimental Psychology: General*, *125*(4), 387–402.

Wong, & Kwong (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 314–333. https://doi.org/10.1006/obhd.2000.2900

Wood, S [Stacy], Poynor, C., & Chartrand, T. (2011). Individual Susceptibility to Priming Effects. *ACR North American Advances*.

Open Science Statement

Almost all empirical studies I was involved in since the beginning of this dissertation in September 2017 have been rigorously preregistered or feature a preregistered direct replication. That is, hypothesis, methods, and analysis plans have been determined *before* data collection. Deviations from the preregistrations are highlighted and discussed.

Data, code, materials, and preregistrations for all studies are available online. Links to the OSF repository are provided with the citations in text. All further material used for this research article is available in the designated OSF project (https://osf.io/9gju4/). I encourage other researchers to make use of the materials and to replicate the studies.

In all studies, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012). Furthermore, I report all studies that I conducted on anchoring effects between September 2017 and August 2020.