RATIONALITY
& COMPETITION
CRC TRR 190

Visual and Social Anchoring in a Framed Online Rating Experiment

Yigit Oezcelik (University of Liverpool)
Michel Tolksdorf (TU Berlin)

Discussion Paper No. 556

December 22, 2025

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

Visual and social anchoring in a framed online rating experiment

Yigit Oezcelik[*], Michel Tolksdorf[†]

Abstract

We conduct an online experiment to assess the effect of the anchoring bias on consumer ratings. We depart from the canonical anchoring literature by implementing non-numerical (visual) anchors in a framed rating task. We compare three anchoring conditions, with either high, low, or socially derived anchors present, against two control conditions – one without anchors and one without framing. Our framing replicates the common observation of overrating. We unveil asymmetric non-numerical anchoring effects that contribute to the explanation of overrating. Both high anchors and socially derived anchors lead to significant overrating compared to the control condition without anchors. The latter finding is driven by instances of high social anchors. The upward rating bias is exacerbated in a social context, where participants exhibit more trust in anchors. In contrast, low anchors and instances of low social anchors have no effect compared to the control condition without anchors. Beyond consumer ratings, our results may have broader implications for online judgment environments, such as surveys, crowdfunding platforms, and other user interfaces that employ visual indicators such as stars, bars, or progress displays.

Keywords: Anchoring bias, consumer judgment, economic experiment, online feedback systems, user interface design

JEL Codes: C91, D80, D91

[*]University of Liverpool, University of Liverpool Management School, Mulberry Court, Mulberry Street, Liverpool L69 7ZH, United Kingdom, Yigit.Oezcelik@liverpool.ac.uk. https://www.liverpool.ac.uk/management/staff/yigit-oezcelik/.

[†]Technische Universität Berlin, H91, Str. des 17. Juni 135, 10623 Berlin, Germany, michel.tolksdorf@tu-berlin.de. https://www.tu.berlin/mikro/team/wissenschaftliche-mitarbeiterinnen/dr-michel-tolksdorf. We thank Christian Bach, Konstantinos Georgalos, Radosveta Ivanova-Stenzel, Tim Jackson, and participants of the ESA World Meetings 2021 and 2023 for useful comments and advice. This work was supported by the Deutsche Forschungsgemeinschaft through the CRC TRR 190 (project number 280092119); and by the Economic and Social Research Council through NWSSDTP.

1 Introduction

Trust among economic agents is crucial for online markets to function. A common way to establish and sustain trust in markets is to facilitate reputation building, e.g., through voluntary online feedback. Nowadays, online ratings are ubiquitous and have become an important part of our everyday lives. When making a purchase, deciding where to eat, or which doctor to visit, many of us consult ratings.

Consumers rely on ratings to make better-informed decisions, as ratings can help them to identify products and services that best match their preferences and needs. Chevalier and Mayzlin (2006) and Dellarocas et al. (2007) show that ratings can significantly affect consumer behaviour and product success in general. Dellarocas (2003) analyses reputation building through a rating system and finds that rating systems can increase market efficiency and foster cooperative outcomes and trust between buyers and sellers. Comparable results are obtained by Chen and Xie (2008), Bohnet and Huck (2004), and Bolton et al. (2004). Moe and Trusov (2011) show that rating dynamics directly influence product sales.

However, ratings can have unintended effects, as even single (bad) transactions can cause reputational externalities for sellers (Nosko and Tadelis, 2015) and bandwagon behavior of buyers occurs due to the social dynamics of online ratings (Moe and Trusov, 2011). Accordingly, research on rating provision has uncovered a J-shaped distribution of ratings translating into some low ratings and many high ratings with little in between these two extremes (Hu et al., 2009). Such inaccurate, upward biased, ratings have been shown to impair consumers' decision-making resulting in suboptimal decisions (Magnani, 2020). Fisher et al. (2018) demonstrate that people tend to rely on a binary heuristic to differentiate between high and low ratings, where an imbalance in favor of high over low ratings drives product purchases. As shown by Givi et al. (2024), consumers are affected by raw scores even when objectively more reliable percentile information is available. While the consequences are known, there is a lack of evidence on what behavioural channels are driving the upward skewness of ratings.

Decades of research in behavioural economics have shown that human decision-making is imperfect and prone to errors and biases. A recurring relevant bias is the anchoring effect, which was first proposed by Slovic and Lichtenstein (1972). This was further elaborated by Tversky and Kahneman (1974), who have shown anchoring happens through irrelevant informational cues substantially affecting behaviour: Excess weight is placed onto a reference point, which is then utilised to anchor decisions, resulting in a failure to fully take into account information that is available subsequently. If ratings were presented in a manner reflective of the J-shaped distribution commonly observed in ratings, this could generate anchoring effects, potentially leading to the upward spiral observed in feedback systems.

Using an incentivised online experiment with payoff-relevant and repeated decisions, our objective is to scrutinise the prevalence and persistence of the anchoring bias in online ratings by isolating the post-purchase and post-consumption rating decision. Our experiment addresses the peculiarities of online rating systems with two main features. First, we consider the perception of the social component of rating systems by varying whether the anchor is given exogenously or derived endogenously with respect to decisions of group members, since anchors derived in a social context increase biased behaviour compared with neutral anchors (Meub and Proeger, 2015). Second, we focus on the visual, i.e., non-numerical, component of ratings, since ratings are often presented as visual stimuli representing higher or lower ratings in the form of filled-out stars, color gradients, happy or frowning smiley faces, et cetera. In physical laboratory settings, it is common practice to remove numerical values (e.g., booth numbers or clocks) except those used as anchors to isolate effects (e.g., in Ivanova-Stenzel and Seres, 2021). However, replicating this practice online is not feasible. Therefore, to maintain experimental control, we employ non-numerical anchors and avoid using any numerical representations in our experiment.

The anchor is presented as a suggested rating, which participants can implement or adjust. In our framed rating task, we present participants either with no anchor, a high anchor (upper bound of the rating scale), a low anchor (lower bound of the rating scale), or a social anchor (average rating of the previous and independent round). All anchors are within the boundaries of the rating scale and, by that, plausible. Sugden et al. (2013) have shown that plausible anchors are more effective than implausible anchors.[1] Additionally, we implement an unframed control condition. This serves as a benchmark, controlling for input errors, and enables us to identify whether the framing of the rating task is effective. Lastly, we explicitly control for the impact of cognitive ability and statistical aptitude, as Hoffart et al. (2019) found that numeracy skills affect rating behavior.

The comparison with the unframed control condition reveals that participants overrate in all framed rating tasks (both with and without anchors present). Concerning high and low anchors, we report asymmetric anchoring effects comparable to Thorsteinson et al. (2008) and Jung et al. (2016). Participants provide upwards biased ratings when they are presented with a high anchor, while the low anchor has no effect. Furthermore, the endogenously derived social anchor affects ratings and is perceived as more relevant. This effect is driven by high socially derived anchors – consistent with our findings on high and low anchors. Overall, our findings help to explain the formation of rating decisions and offer insights to design less error-prone rating platforms that avoid anchoring.

[1] An issue of using suggested ratings as anchors is that (plausible) direct choices of suggestions could be driven by default effects and anchoring effects. To cleanly identify anchoring effects, we disregard "default" observations in our main analysis.

We contribute to the literature in two ways. First, we advance the understanding of (social) anchoring effects in rating settings by providing controlled experimental evidence. Second, to our knowledge, we are the first to demonstrate the impact of non-numerical, visual, anchors on economically framed decisions. Although our findings are grounded in the context of consumer ratings, they may extend more broadly to other online judgment settings, including, for example, survey responses or crowdfunding platforms that feature visual progress indicators such as progress bars.

2 Literature Review

The anchoring bias has been shown to be pervasive and robust in many situations, from individual consequential economic decisions such as credit card minimum repayments (McHugh and Ranyard, 2016), asset value assessments (Ünveren and Baycar, 2019), real estate evaluations (Northcraft and Neale, 1987), "pay what you want" price choices (Jung et al., 2016) to strategic interactions (Ivanova-Stenzel and Seres, 2021, 2022), auctions (Chui et al., 2022) and bargaining (Kimbrough et al., 2021) (for a comprehensive literature review on the anchoring effect, see Furnham and Boo, 2011).[2]

However, the evidence on pervasiveness and robustness of anchoring effects in willingness-to-accept/pay, originally depicted by Ariely et al. (2003), is mixed. Sugden et al. (2013) find that anchors need to be plausible and that effects are stronger on willingness-to-pay compared with willingness-to-accept. Several experimental studies report weak anchoring effects on willingness-to-pay/accept and provide evidence for a lack of robustness of prior findings (see Alevy et al., 2015; Fudenberg et al., 2012; Maniadis et al., 2014). On the other hand, Yoon and Fong (2019) and Yoon et al. (2019) argue that anchoring effects on willingness-to-pay are persistent over a long-term, respectively robust to changes in experimental procedure, setting, and participants. As ratings are one underlying factor in the formation of the willingness-to-pay for goods, our study is closely aligned with this literature.

The aforementioned studies use exogenous anchors in autonomous and individual decision-making environments. However, many economic decisions are made in social settings. Bischoff and Egbert (2013) highlight how social information can affect individual decisions. Besides incorporating monetary incentives and learning through feedback, it is important to consider the social context when scrutinising anchoring effects. Observing other individuals' behaviour and the information they are sharing can enhance learning effects. In fact, there is

[2] Four classes of explanation have been offered: i.) underadjustment (Tversky and Kahneman, 1974; Epley and Gilovich, 2001), ii.) numerical priming (Jacowitz and Kahneman, 1995) , iii.) confirmatory hypothesis testing (Chapman and Johnson, 1999) and iv.) scale distortion theory (Frederick and Mochon, 2012).

some evidence of social anchoring. Phillips and Menkhaus (2010) investigate the effect of an endogenously derived anchor on willingness-to-pay/accept decisions in an auction environment. Here, the average price of the previous round serves as an anchor. The authors find evidence of significant anchoring effects. Using an economically incentivised estimation task and allowing for learning through feedback, Meub and Proeger (2015) also present robust social anchoring effects. In their case, the average estimate of the previous round serves as the anchor. They compare the socially derived anchor to an exogenous anchor and show that the social context increases the anchoring bias.

Ratings are provided in social environments. While Li et al. (2020) argue that allowing sellers to buy feedback can be beneficial, there usually is no direct incentive to provide ratings. Evidence on what drives the formation of rating decisions in social domains points toward an upward bias. Bolton et al. (2019) cast doubt on the informativeness of ratings, as uncertainty fosters leniency resulting in upward inflated ratings. Similarly, Coker (2012) provides evidence of asymmetrical affective perseverance when forming attitudes based on prior information. Consumers overshoot their judgments when positive information is replaced with negative information but not vice versa. This means positive reviews generate positive attitudes, even if negative reviews follow later and consumers know previous information was erroneous. There is some evidence on behaviour that is explicitly indicative of the anchoring bias in case of performance ratings. Thorsteinson et al. (2008) used field and laboratory studies to scrutinise anchoring effects in performance judgements, finding asymmetrical anchoring effects where the low anchors have a weaker effect than high anchors.

Most research that investigates the anchoring bias, including, to our knowledge, all studies on social anchors and performance ratings, rely on numerical anchors. In this paper, we diverge from this literature by employing non-numerical, visual anchors. In rating environments, a lot of information is displayed not only numerically but (sometimes exclusively) visually via coloured stars or bars, smileys, emojis (thumbs up/down), or color-coded indicators. For instance, on Amazon and Uber continuous five-star rating systems are used. The user must colour the number of stars that they want to provide. Seeing the 5 stars template from the beginning might already serve as a high anchor and create a sense of largeness. Earlier literature discovered evidence on other, but comparable, non-standard ways of anchoring. LeBoeuf and Shafir (2006) report physical anchoring effects, using various stimuli such as length, weight, or volume. Oppenheimer et al. (2008) provide experimental evidence for cross-modal effects of physical anchors. The authors explain the effect via magnitude priming which is the creation of a perception of largeness (or smallness). However, these studies lack monetary incentives, the possibility for learning, and the application to an economic context.

Our study builds on these insights by transferring the logic of physical and cross-modal anchoring into an economic decision-making setting. Unlike the earlier laboratory studies, our experiment incorporates monetary incentives, repeated decisions, and a framed online rating environment. In this sense, our study can be viewed as an application and extension of the anchoring mechanisms identified by LeBoeuf and Shafir (2006) and Oppenheimer et al. (2008) to an economically framed decision-making context.

3 The Experiment

We report an experiment, using a between-subject design with five experimental conditions: No Framing, No Anchor, High Anchor, Low Anchor, and Social Anchor (see Table 1).

Condition	Task	Anchor (high/low)	Context	Effect(s)
No Framing	Unframed	No	Nonsocial	–
No Anchor	Rating	No	Nonsocial	Framing
High Anchor	Rating	Yes (high)	Nonsocial	Framing, Anchoring
Low Anchor	Rating	Yes (low)	Nonsocial	Framing, Anchoring
Social Anchor	Rating	Yes (high/low)	Social	Framing, Anchoring, Trust

Table 1: Overview of the five experimental conditions.

We use a framed individual decision-making task where participants provide quality ratings. Participants are presented with a slider that represents the overall range of quality. On this slider, a quality interval is represented by a portion marked by two bars, spanning approximately one-fourth of the slider's length. Participants are instructed to estimate the true quality within this interval by adjusting the slider to the corresponding position. The exact position of the true quality in the quality interval is unknown to the participants. Participants only know that the interval always contains the true quality, with each value within the interval having an equal probability of being randomly drawn as the true quality.[3]

The uncertainty of the true quality within a quality interval mimics the notion that people often receive noisy quality signals upon which they make rating decisions (Bolton et al., 2019). We present participants with an abstract setting in which they do not rate any specific product for two purposes. First, it eliminates the possibility of magnitude priming which could be product-specific (LeBoeuf and Shafir, 2006; Oppenheimer et al., 2008). Second, it also eliminates the impact of participants' pre-existing attitudes which could carry over asymmetrically and, therefore, introduce an upward bias (Coker, 2012).

[3]The full translated and original instructions can be found in 5.4.2 in the Appendix. The instructions include depictions of the slider interface.

In three out of four rating task conditions there is an anchor present while the No Anchor condition serves as a control to gauge the effect of the presented anchors. In the No Framing condition everything exactly corresponds to the No Anchor condition, except that participants receive neutrally framed instructions. In particular, this means that the task is not framed as a rating task. Instead, participants are explicitly instructed to move the slider's handle to an invisible target value which is always positioned in the middle of the interval. Thus, the unframed task is meant to check how accurately participants can utilise the slider's handle if they intend to adjust it to the middle.

The closer the participants' estimates are to the true quality in the rating task or the target value in the unframed task, the higher their payoff. In all conditions the respective task is repeated for twelve rounds without feedback between rounds. There are no numerical values displayed on the slider, the quality interval, or anywhere else on the experimental screen. We utilise pixels (`px`) to measure the provided ratings and to determine payoffs. The CSS unit `px` is usually understood as the smallest unit of measurement in CSS applications. For payoffs, one `px` equals one Experimental Currency Unit (ECU) and 100 ECU equal 1 EUR. This in turn means that each minuscule movement of the slider is payoff relevant for the participants. As participants never learn the true quality, the payoff maximising strategy is to choose the expected quality value, i.e., the middle of the quality interval, to minimise (expected) deviation from the true quality. In other words, the (expected) payoff maximising rating is $r_t^* = (q_t^H - q_t^L)/2$ (henceforth, "optimal rating") in round $t \in \{\text{one}, \ldots, \text{twelve}\}$ where q^H is the upper bound of the quality range and q^L is the lower bound of the quality range (see 5.1 in the Appendix for the formal derivation of the optimal rating).

The positions and length of the displayed quality intervals were randomly and independently drawn for all rounds using a common seed number so that they are kept the same across all sessions and conditions to ensure comparability. Further, to ensure the irrelevance of anchor values in the Low and High Anchor conditions, both the bottom and top 50`px` of the slider range were excluded from the draw. Table 2 shows the draws of the lower quality bounds q_t^L, upper quality bounds q_t^H, quality interval lengths $q_t^H - q_t^L$, true qualities q_t and optimal ratings r_t^* for all twelve rounds t. Subsequently, the maximum (1000`px`) and minimum (0`px`) values are always outside the quality range and are used as anchor values in the High Anchor and Low Anchor conditions, respectively.

In the High and Low Anchor conditions, the participants can either choose the presented anchor as their rating or choose to adjust the rating. If participants choose to adjust the rating, they must reinitialise their rating by clicking anywhere on the slider. These two anchor types present exogenous and irrelevant informational cues as the quality interval

Round	1	2	3	4	5	6	7	8	9	10	11	12
q_t^L	497	603	316	464	143	73	474	455	422	93	78	467
q_t^H	796	846	490	655	423	366	750	746	578	392	372	675
$q_t^H - q_t^L$	299	243	174	191	280	293	276	291	156	299	294	208
q_t	748	666	391	628	386	331	608	690	464	208	249	577
r_t^*	646	724	403	560	283	220	612	600	500	242	225	571

Table 2: Quality bounds q_t^L and q_t^H, interval length $q_t^H - q_t^L$, true quality q_t and optimal rating r_t^* by round t.

never contains the minimum or maximum of the slider.[4] In contrast to the High and Low Anchor conditions, the anchor is derived endogenously in the Social Anchor condition. The participants are divided into groups of five that remain the same throughout the experiment. At the beginning of every round, starting from round two on, the group's average rating from the previous period serves as the anchor and is shown to the participants. The participants can either select the socially derived anchor or can choose to adjust their rating. For instance, if the participants are in round two, they see the average rating across all group members from round one as the anchor.

The socially derived anchor does not contain any additional informational value as the quality interval is independently drawn in each round. However, in some instances, the social anchor is contained in the quality interval. We will discuss the implications of this further below in the results section. In the No Anchor condition, there is no anchor present in any form. Hence, in this case participants always have to initialise the slider by clicking on it and can readjust their rating by dragging the slider's handle. The No Framing condition is like the No Anchor condition, but in absence of an explicit framing, as explained before.

In addition to the rating task, respectively unframed task, we collected several control measures in a post-experimental questionnaire. We ran a standard cognitive reflection test (CRT). Additionally, we ran a questionnaire to get an indication of statistical aptitude that consists of three questions. In the first two questions, we asked participants to state whether they have any prior knowledge of statistics and whether they know what an expected value is. In the third question, participants were asked to work out the expected value for a dice roll problem. Only in the anchoring conditions, we additionally measure trust in the anchor. We operationalise trust as participants' perceived relevance of the anchor, measured through self-reported ratings of how informative, helpful, and influential they found the anchor on 7-point Likert scales. Our trust measure is the sum of the three responses normalised to

[4]See Chapter 5.2 in the Appendix for an exemplary decision screen of round one in the High Anchor condition.

$[0, 1]$.[5] Before the experiment started, we provided the participants with detailed instructions and control questions to ensure comprehension of the task. Depending on the experimental condition, participants then played the twelve rounds of the rating task or unframed task. Afterward, participants answered the post-experimental questionnaire, including the CRT test, statistical aptitude questions, and, in the case of anchor conditions, questions regarding trust towards the presented anchors.

3.1 Participants

Experiments were programmed using oTree (Chen et al., 2016) and conducted online in April, May, and October 2021 using the ORSEE database and student participant pool of the joint laboratory of WZB Berlin and TU Berlin (Greiner, 2015). Each participant only participated in one experimental condition and was unaware of the other conditions. On average, participants earned 14 EUR, including a show-up fee of 5 EUR. Sessions lasted for 30 minutes. In the end, one round was selected at random to determine the final payoff. We conducted 5 sessions per experimental condition. In total, 246 participants took part in the experiment: 50 in the Social Anchor condition and 49 in each of the other conditions.[6]

An a-priori power analysis was not feasible due to the novelty of our experimental design. To our knowledge our outcome of interest (deviation of slider handle positions in `px`) has not been measured before in a comparable context. Moreover, since we collected panel data, proper power analysis is only possible using a simulation-based approach, drawing from an existing data set (Burlig et al., 2020). Hence, we opted to determine our sample size based on the previous literature on social anchoring, aiming for a comparable number of independent observations, i.e., 50 participants per condition (cf. Meub and Proeger, 2015, 2016, who invited 35 to 58 participants per condition). In terms of observations per participant, we relied on a simple heuristic. Every screen of our experiment was supposed to be devoid of any form of numbers or numerical representations. Since we ran the experiment in German with a majority of German participants, we featured 12 rounds. It is a convention in German writing to represent only numbers up to 12 as written words and numbers thereafter numerically. Therefore, participants were reminded that they are in "round [one, . . . , twelve] of twelve" solely in written words. While our number of rounds is slightly lower than the 15 rounds used in Meub and Proeger (2015, 2016), Burlig et al. (2020) argue that benefits to power when increasing panel length erode quickly once surpassing short panel lengths of 5 or less.

[5]See 5.3 in the Appendix for the post-experimental questionnaire.

[6]The procedure of our experiment is comparable to laboratory experiments in terms of participant pool, duration, and payoffs.

3.2 Hypotheses

In our experimental setup, the anchors are designed to be useless and to contain no relevant informational value. This means, if participants are not anchored, their (expected) payoff maximising choice is to select the middle of the quality interval to minimise deviations from the expected quality and maximise expected payoffs. Due to the prevalence of the anchoring effect in many economic environments, we expect anchoring effects in the rating task.

We hypothesise that ratings will be biased towards the thresholds in the High and Low Anchor conditions, respectively.

Hypothesis 1a: *Ratings are biased upwards in the High Anchor condition.*
Hypothesis 1b: *Ratings are biased downwards in the Low Anchor condition.*

Participants make repeated rating decisions over twelve rounds. By this we can investigate the persistence of anchoring effects, as participants might be less prone to the anchoring bias in the later rounds of the experiment.

In the Social Anchor condition, the anchor is derived in a social context by averaging the previously provided ratings within a group. Hence, depending on the ratings of the previous round this might give rise to both high and low socially derived anchors. Like before, we expect both types of social anchors to bias ratings.

Hypothesis 2: *Ratings are biased towards the group anchor in the Social Anchor condition.*

As a social context can increase the perceived relevance and informativeness of the anchor (De Wilde et al., 2018), we expect that the anchor will be perceived as more trustworthy in the Social Anchor condition compared to the other anchor conditions. Furthermore, we expect that the anchoring bias is more pronounced for individuals who place greater trust in the socially derived anchor.

Hypothesis 3a: *The anchoring bias is more pronounced when the anchor is perceived as more relevant.*
Hypothesis 3b: *Socially derived anchors are perceived as more relevant than high and low anchors.*

The conceptual framework in Figure 1 summarises our hypotheses and how we isolate the expected anchoring effects from other confounding factors with our experimental design. We control for mere input errors by means of the unframed task, for cognitive ability and statistical aptitude by means of a questionnaire, and for any other confounding factors by keeping the rating task fixed between the anchoring and No Anchor conditions. The comparison between the rating task conditions and the unframed task ensures that the framing of the experiment is effective and serves as a data-based comparison measuring differences between observed ratings and the expected (true) quality. We expect that ratings are biased upwards in the high anchor condition (H1a), downwards in the low anchor condition (H1b), and towards the high, respectively low, socially derived anchor in the social anchor condition (H2). We expect that trust moderates the anchoring effect (H3a) and is itself amplified by the social context of the social anchor condition (H3b).

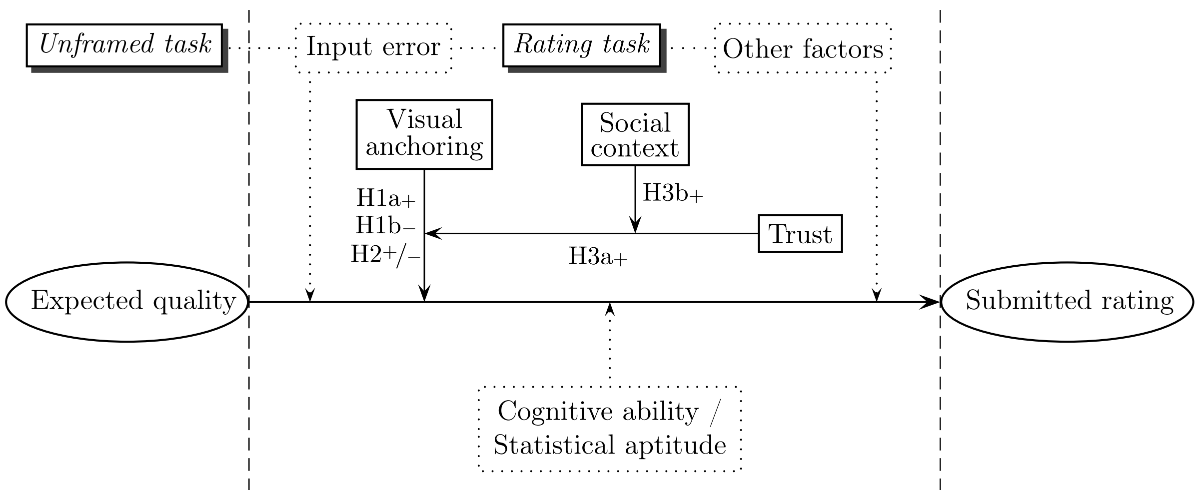


Figure 1: Conceptual framework.

3.3 Experimental Results

In the following, we explore our hypotheses using the experimental results. We primarily refer to the effects of experimental conditions and control variables on either the absolute rating r_{it} or the normalised rating $\tilde{r}_{it} = (r_{it} - r_t^*)/(q_t^H - q_t^L)$. This normalisation expresses changes as percentages of the quality interval length. Because the rational rating r_t^* is the midpoint between q_t^H and q_t^L, the normalised scale ranges from -0.5 (corresponding to q_t^L) to $+0.5$ (corresponding to q_t^H).[7]

For our main analysis, we exclude those observations where participants directly chose the presented anchor instead of adjusting the rating as such decisions have two caveats. First, participants may have willingly defaulted to the presented anchor, which could turn eventual

[7]For example, an effect size of $+0.05$ (or -0.05) would correspond to a 10 percentage point increase (or decrease) within the quality interval relative to the rational rating r_t^*.

findings on anchoring effects indiscernible from default effects. Second, even when participants made this decision by mistake and wanted to change it, it was irreversible for that given round. We drop 122 such observations from the rating task conditions. Additionally, we disregard one clear outlier in No Framing where a participant chose a value outside the quality range and 297px from the target value, while all other 587 observations in No Framing are within 21px of the target value. This leaves us with a total of 2,829 observations by 246 participants. Nearly all participants in the framed conditions made choices that deviate substantially from the choices observed in the No Framing control condition.[8] In line with Jung et al. (2016), we consider this a conservative approach to estimate anchoring effects. All significant results of our main analyses are robust to not dropping any observations. However, this would introduce default effects leading to the dubious finding of an initial negative impact of low anchors and exaggerated effect sizes (see 5.4 in the Appendix for robustness checks on all regressions utilising the full data set).

Experimental condition	Mean	Std. Dev.	Min	Max
No Framing	0.001	0.015	-0.052	0.070
No Anchor	0.027	0.207	-0.667	0.829
High Anchor	0.058	0.233	-0.873	1.994
Low Anchor	0.049	0.248	-1.317	1.526
Social Anchor	0.069	0.300	-1.053	1.609

Table 3: Summary statistics of normalised ratings by condition.

In Table 3, we show the summary statistics of the normalised ratings across all experimental conditions. In Figure 2, we show the normalised ratings over rounds. Taken together, we observe that the Social Anchor condition diverts ratings the most in comparison with the No Framing and No Anchor conditions due to spikes in particular rounds. The High (Low) Anchor condition diverts ratings in the early (late) rounds compared with the No Framing and No Anchor conditions. All anchors seem to inflate ratings, with the highest impact of the high and social anchors. Surprisingly, even in the Low Anchor condition we observe on average higher ratings compared with the No Anchor condition.

To test for treatment effects of the anchoring conditions compared with the control conditions on normalised rating $\tilde{r}_{it}$ of participant i at time t, we employ multi-level models of the following form:

$$\tilde{r}_{it} = \mu + \mathbf{d}_{\mathbf{i}}'\boldsymbol{\theta} + \eta h_t + h_t\mathbf{d}_{\mathbf{i}}'\boldsymbol{\psi} + \mathbf{x}_{\mathbf{i}}'\boldsymbol{\phi} + u_i + \epsilon_{it}. \quad (1)$$

[8] Only 5 participants (3 in Low Anchor, 2 in High Anchor) in the framed conditions made all decisions within 12px of the rational rating. No other participant in the framed conditions made all decisions within a widened bound of 36px. All results presented throughout are robust to dropping these participants.

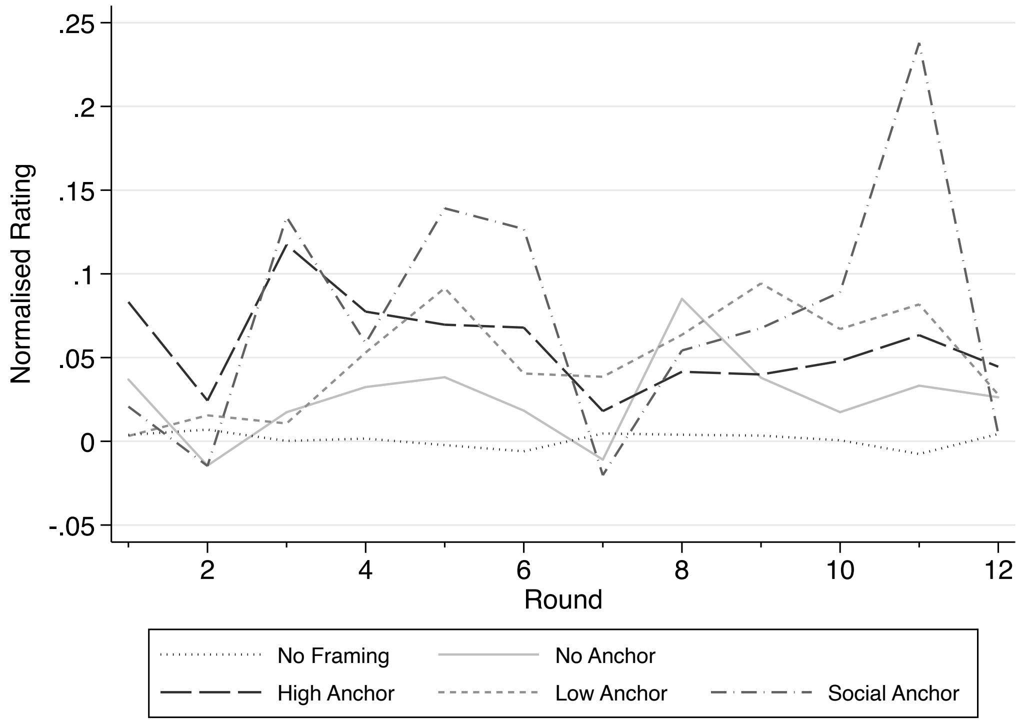


Figure 2: Average normalised ratings over rounds by experimental condition.

In this model, $\mathbf{d_i}$ denotes the vector of dummies indicating the experimental conditions with $\boldsymbol{\theta}$ indicating the corresponding coefficient vector. In the advanced specifications h_t denotes a dummy variable indicating whether observations are from rounds seven to twelve with η capturing its direct effect, and $\boldsymbol{\psi}$ capturing interaction effects with condition dummies. The parameter vector $\boldsymbol{\phi}$ captures the effects of controls in form of CRT and statistical aptitude scores contained in $\mathbf{x_i}$. The nonrandom scalar intercept is μ. Lastly, the composite error term consists of the subject-level random intercepts u_i and the conventional error term ϵ_{it}.[9]

In Table 4, we show the main effects of the regressions on whether anchors in any experimental condition have an impact on the normalised rating compared to the No Framing and No Anchor conditions. In specifications (1) and (2), we compare the anchoring conditions to the No Framing control. Following specification (1), we observe significant overrating in all rating task conditions compared to the unframed task in the No Framing control, indicating that our framing is effective. When employing a dummy variable indicating the second half

[9]For repeated observations without feedback, the literature often employs OLS regressions with clustered standard errors. Cheah (2009) shows by means of simulations that multilevel models yield more reliable and conservative inferences than OLS regressions. We report the intraclass correlations (ICC) and follow the classification of Hox et al. (2017). We observe significant ICC values around 0.05 to 0.1, indicating small to moderate intraclass correlation in line with most social science studies. As a robustness check, we report the results of OLS regressions with clustered standard errors in Chapter 5.4.2 in the Appendix.

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
High Anchor	0.058***	0.074***	0.035*	0.056**
	(0.013)	(0.017)	(0.017)	(0.021)
Low Anchor	0.046***	0.033*	0.023	0.015
	(0.010)	(0.016)	(0.015)	(0.020)
Social Anchor	0.066***	0.076***	0.043*	0.057*
	(0.017)	(0.021)	(0.021)	(0.025)
No Anchor	0.023*	0.019		
	(0.011)	(0.013)		
Second half		0.001		0.010
		(0.001)		(0.014)
High Anchor × Second half		-0.032†		-0.041†
		(0.018)		(0.023)
Low Anchor × Second half		0.026		0.017
		(0.020)		(0.025)
Social Anchor × Second half		-0.019		-0.028
		(0.026)		(0.030)
No Anchor × Second half		0.009		
		(0.014)		
Constant	0.008	0.008	0.033**	0.028†
	(0.007)	(0.007)	(0.012)	(0.015)
Intraclass correlation	0.053***	0.054***	0.052***	0.053***
Observations	2829	2829	2242	2242
Number of subjects	246	246	197	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by multi-level model with subject-level random-effects. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

Table 4: Treatment effects on normalised rating.

of the experiment (rounds seven to twelve) and interaction terms in specification (2), we find that only the effects of the anchor conditions carry through.

Specifications (3) and (4) mirror the approach from before relative to the No Anchor condition, i.e., only for the rating task conditions. In specification (4), we again include a dummy variable indicating the second half and its interactions with the conditions. We find that both under the High Anchor condition and the Social Anchor condition normalised ratings are significantly greater than under the No Anchor condition, whereas we do not find a significant effect for the Low Anchor condition. Contrary to our hypothesis, the sign of the effect of the Low Anchor condition is even positive. The effect of high anchors declines over time as indicated by the marginally significant interaction High Anchor × Second half.

Result 1: *The High Anchor condition has a positive effect on ratings. The Low Anchor condition has no negative effect on ratings.*

In the following, we explore the role of social anchors on rating behaviour. As the socially derived anchor can be high or low, we proceed in two steps. First, we check whether the socially derived anchor is predictive of the observed rating when controlling for the optimal rating r^*. Second, we check whether the bias differs between high and low social anchors. We classify social anchors according to two types: High social anchor ($\bar{r}_{t-1} > r_t^*$), and low social anchor ($\bar{r}_{t-1} < r_t^*$). The following multi-level random-slopes model explains rating r_{it} of participant i at time t in terms of the optimal rating r_t^* and the anchoring deviation $r_t^* - \bar{r}_{t-1}$ at time t:

$$r_{it} = \alpha r_t^* + \beta(r_t^* - \bar{r}_{t-1}) + \mathbf{x_i'}\boldsymbol{\phi} + v_i r_t^* + u_i(r_t^* - \bar{r}_{t-1}) + \epsilon_{it}. \tag{2}$$

For fully payoff maximising participants, we would expect $\alpha = 1$ and $\beta = 0$. That is, the rating is fully described by the optimal rating r^*. Anchoring deviations, i.e., biased ratings towards the anchors, are present when $\beta < 0$. Again, the parameter vector $\boldsymbol{\phi}$ captures the effects of controls in form of CRT and statistical aptitude scores contained in $\mathbf{x_i}$. Lastly, the composite error term consists of v_i and u_i denoting the subject-specific random-slopes, while ϵ_{it} denotes the conventional error term. First, we estimate this model for all observations to identify the general anchoring deviation in the Social Anchor condition. Subsequently, we divide our observation pool according to our classifications into observations where a high social anchor ($\bar{r}_{t-1} > r_t^*$) is present, and observations where a low social anchor ($\bar{r}_{t-1} < r_t^*$) is present.

We show the main effects of the regression results of the aforementioned model in Table 5. In specification (1), we find that α is not significantly different from one via a post-estimation Wald test at the 95% confidence level. Further, we can confirm the anchoring deviation by a significant $\beta < 0$. By splitting the observations into low and high social anchors we observe that both these effects are still present in (3), where we restrict our attention to high social anchors, while we observe no anchoring deviation in (2), where we consider only low social anchors. We conclude that high social anchors distort ratings upwards, whereas low social anchors induce participants to give close to optimal ratings. Specification (4) mirrors the same approach for the High Anchor condition which corresponds to fixing $\bar{r}_{t-1} = 1000$. The result indicates that the anchor deviation depicted for high social anchors is present there as well, albeit with a lower effect size.[10]

We depict the main intuition of the described relation between the social anchor and the normalised ratings in Figure 3. The observed normalised ratings follow the path of high

[10] We cannot apply the same approach for the Low Anchor condition due to collinearity of anchor deviation and optimal rating, since in that case $\bar{r}_{t-1} = 0$.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.027***	1.008***	1.029***	1.011***
	(0.0153)	(0.0198)	(0.0228)	(0.0112)
Deviation ($r_t^* - \bar{r}_{t-1}$)	-0.0766***	-0.0190	-0.0900**	-0.0270*
	(0.0168)	(0.0424)	(0.0316)	(0.0112)
χ^2-statistic ($\alpha = 1$)	3.20	0.15	1.66	0.91
Prob. > χ^2	0.074	0.695	0.198	0.339
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	462	193	269	576
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by multi-level model with subject-level random-slopes. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

Table 5: Regression results of equation (2).

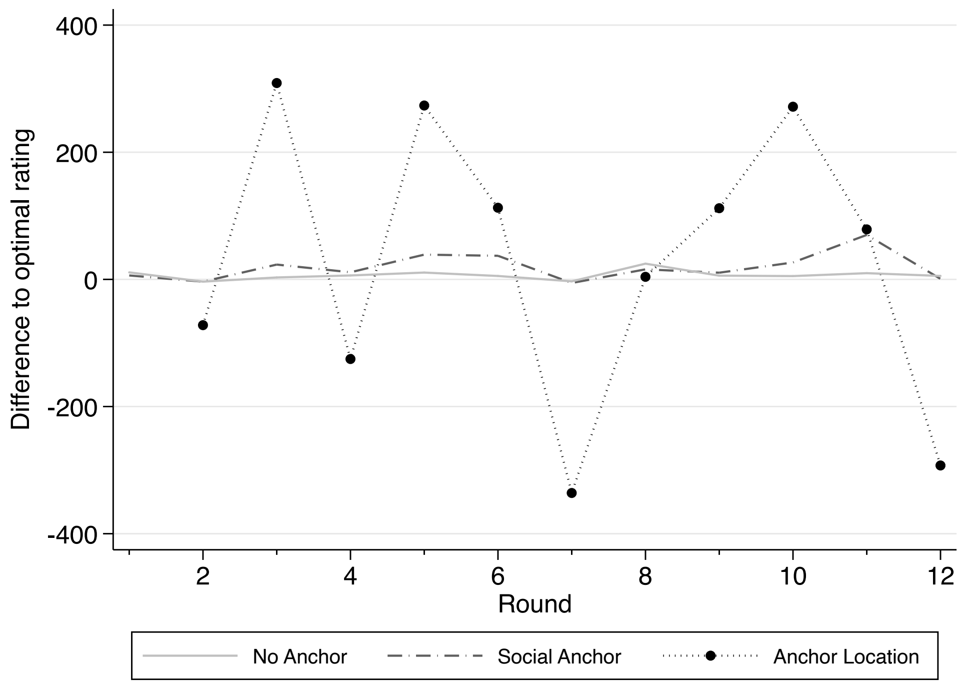


Figure 3: Relation of social anchor location and mean ratings.

social anchors. The upwards trends do not faint for subsequent high social anchors. Low social anchors pull ratings down close to the level of the No Anchor condition. We have to note that we have only a few instances of two subsequent high social anchors and no such cases for low social anchors.

Round	1	2	3	4	5	6
SA classification	N/A	Low	High	Low	High	High
SA norm. rating	6.200	-3.521	23.319	11.190	38.959	37.147
NA norm. rating	11.040	-3.581	3.020	6.188	10.714	5.367
Δ norm. rating	-4.840	0.060	20.299	5.002	28.245	31.780
Round	7	8	9	10	11	12
SA classification	Low	Mixed	High	High	High	Low
SA norm. rating	-5.580	15.783	10.542	26.617	69.967	0.878
NA norm. rating	-3.020	24.755	5.918	5.204	9.771	5.469
Δ norm. rating	-2.560	-8.972	4.624	21.413	60.196	-4.591

Table 6: Classification of Social Anchor (SA) types and calculation of difference between mean normalised ratings in Social Anchor condition and No Anchoring (NA) condition.

To support these results, we classify rounds by the type of social anchor present and compare the effect of socially derived anchors in these rounds to the High, Low, and No Anchor conditions. We "mark" those rounds where we observe only high anchors, such that for all observations $\bar{r}_{t-1} > r_t^*$, as "High". Likewise, we depict those as "Low" where for all observations $\bar{r}_{t-1} < r_t^*$. If neither condition applies, we depict a round as "Mixed". In Table 6, we show that by this we classify rounds 3, 5, 6, 9, 10 and 11 as High, rounds 2, 4, 7 and 12 as Low. Only round 8 is depicted as Mixed, with both $\bar{r}_{t-1} > r_t^*$ and $\bar{r}_{t-1} < r_t^*$ for some observations, respectively. Out of 50 observations, 27 were dropped in round 8 since participants chose a rating equal to the group anchor. Round 1 has no classification, as there is no anchor present. Further, we calculate the difference of normalised ratings between the Social Anchor and No Anchor conditions as an indication of the extent of the anchoring deviations.

We run regressions using multi-level random-effects models as in (1) where we replaced the second half dummy h_t with the dummy m_t indicating observations from "marked rounds" (with a high social anchor present):

$$\tilde{r}_{it} = \mu + \mathbf{d_i}'\boldsymbol{\theta} + \eta m_t + m_t \mathbf{d_i}'\boldsymbol{\psi} + \mathbf{x_i}'\boldsymbol{\phi} + u_i + \epsilon_{it}. \tag{3}$$

The results are shown in Table 7. In specifications (1) and (3), we show that there is a direct impact of the social anchor in the marked rounds compared to the No Framing condition in (1), respectively the No Anchor condition in (3), as captured by the significant interaction effect. There is neither a direct effect for the social anchor nor for the marked rounds, indicating that high social anchors drive the results of socially derived anchors. To ensure that there are no other impacts in these specific rounds, we show in specifications (2) and (4) that there are no significant interactions of the Low Anchor and High Anchor conditions in the

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
Marked Rounds	-0.006***	-0.006***	0.001	0.001
	(0.001)	(0.001)	(0.017)	(0.017)
Social Anchor	0.007	0.008	-0.012	-0.012
	(0.022)	(0.022)	(0.025)	(0.025)
High Anchor		0.046**		0.026
		(0.015)		(0.019)
Low Anchor		0.028†		0.009
		(0.016)		(0.019)
No Anchor		0.020†		
		(0.011)		
Social Anchor × Marked Rounds	0.117***	0.118***	0.110**	0.110**
	(0.030)	(0.030)	(0.035)	(0.035)
High Anchor × Marked Rounds		0.024		0.017
		(0.022)		(0.028)
Low Anchor × Marked Rounds		0.037		0.029
		(0.026)		(0.032)
No Anchor × Marked Rounds		0.007		
		(0.017)		
Constant	0.015†	0.011†	0.038*	0.033**
	(0.009)	(0.007)	(0.015)	(0.013)
Intraclass correlation	0.094***	0.055***	0.072***	0.054***
Observations	1099	2829	1097	2242
Number of subjects	99	246	99	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by multi-level model with subject-level random-effects. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. * *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

Table 7: Impact of high socially derived anchors.

marked rounds. We conclude that our depiction captures the impact of social anchors in that, specifically, high social anchors affect ratings. Hence, we can partially confirm Hypothesis 2.

Result 2: *A high social anchor distorts ratings upwards, while a low social anchor has no distorting effect.*

Next, we explore the role of trust in the anchoring conditions. We derive a normalised trust score which is bounded by 0 and 1. The trust score is the fraction of the sum of numerical responses provided to the three trust-related questions divided by the maximum trust value of 18. The maximum value is 18 since we used a 7-point Likert scale, coded 0 to 6, in each question. In Figure 4, we show kernel density estimations of the trust score. There is a clear indication of higher and more diversified trust scores in the Social Anchor condition. Both in the High Anchor and Low Anchor conditions, participants more often opted for the lowest possible scores, with 33/49 in the High Anchor condition and 28/49 in the Low Anchor condition, compared to 7/50 in the Social Anchor condition.

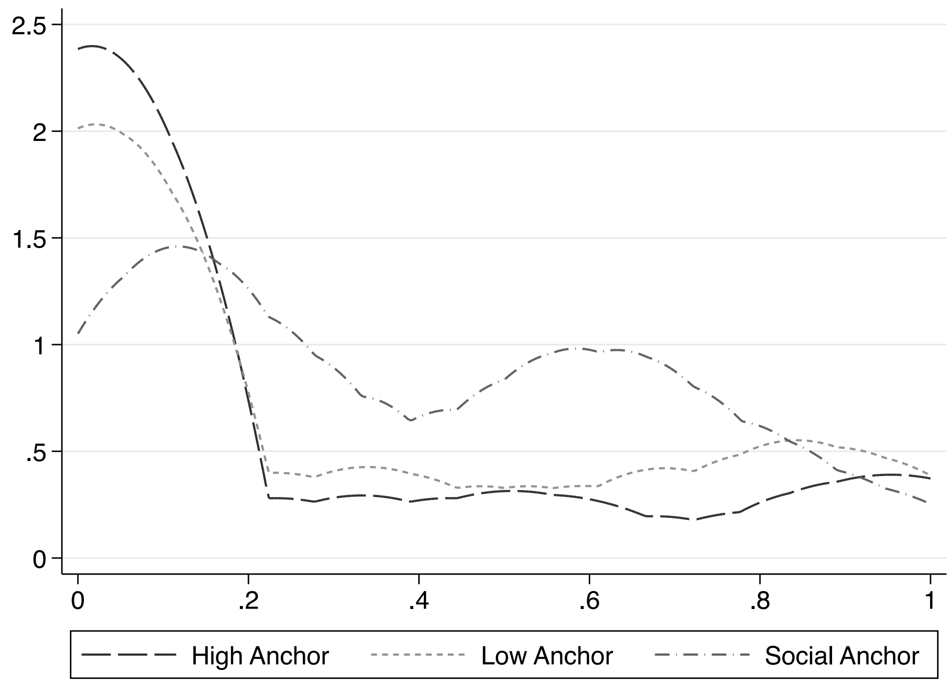


Figure 4: Kernel density estimation of trust score for anchoring conditions.

Trust is significantly higher in Social Anchor (M = .374, SD = .043) compared to the High Anchor condition (M = .192, SD = .048), $t(97) = 2.853, p = .006$ and marginally significantly higher compared to the Low Anchor condition (M = .253, SD = .051), $t(97) = 1.842, p = .068$, based on two-sided t-tests. There is no significant difference between the High Anchor and Low Anchor conditions based on a two-sided t-test, $t(96) = 0.878, p = .382$. All results also hold when using two-sided Wilcoxon rank-sum tests.

Next, we want to determine whether trust in the social anchors drives the anchoring deviation. We extend the multi-level random-slopes model in (2) by an interaction of the normalised individual trust score s_i with the anchoring deviation $(r_t^* - \bar{r}_{t-1})$ and the corresponding subject-specific random-slopes w_i:

$$r_{it} = \alpha r_t^* + \beta(r_t^* - \bar{r}_{t-1}) + \gamma s_i(r_t^* - \bar{r}_{t-1}) + \mathbf{x_i'}\boldsymbol{\phi} + v_i r_t^* + u_i(r_t^* - \bar{r}_{t-1}) + w_i s_i(r_t^* - \bar{r}_{t-1}) + \epsilon_{it}. \quad (4)$$

Like before, we would expect $\alpha = 1$, $\beta = 0$, and $\gamma = 0$ when participants act payoff maximising. If participants are biased by the group anchor irrespective of trust, we would expect $\beta < 0$ and $\gamma = 0$. However, if this bias is exacerbated by trust, we should expect $\beta < 0$, $\gamma < 0$, and $\beta + \gamma < 0$.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.027***	1.009***	1.027***	1.011***
	(0.0153)	(0.0195)	(0.0240)	(0.0113)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.0356	0.00197	-0.0305	-0.0232*
	(0.0258)	(0.0515)	(0.0406)	(0.0113)
Deviation × trust $(r_t^* - \bar{r}_{t-1})s_i$	-0.110†	-0.0570	-0.154†	-0.0222
	(0.0588)	(0.0636)	(0.0912)	(0.0185)
χ^2-statistic $(\alpha = 1)$	3.05	0.19	1.26	0.96
Prob. > χ^2	0.081	0.660	0.262	0.328
χ^2-statistic $(\beta + \gamma = 0)$	12.20***	1.03	7.12**	5.23*
Prob. > χ^2	0.0005	0.3102	0.0076	0.0222
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	462	193	269	576
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by multilevel model with subject-level random-slopes. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

Table 8: Regression results of equation (4).

In Table 8, we present the results of our estimation. Introducing the trust score s_i reveals that we cannot confirm a significant $\beta < 0$, although we do find an indication for $\gamma < 0$ with marginal significance, both for all observations ($p = 0.061$) and for only high social anchors ($p = 0.090$). In particular, we find that the joint test $\beta + \gamma < 0$ is significant. This indicates that the anchoring deviation depicted previously is largely, but not solely, driven by those who place high trust in the signal. Again, we mirror this approach for the High Anchor condition. In that case, including trust as a moderator barely absorbs the anchor deviation-effect and the interaction of anchor deviation and trust is insignificant ($p = 0.231$). Hence, in the High Anchor condition the significance of the joint test $\beta + \gamma < 0$ is mainly driven by the anchoring deviation. In summary, these observations support Hypothesis 3.

Result 3: *The anchor is perceived as more relevant in the Social Anchor condition compared to the High Anchor and Low Anchor conditions. Trust moderates the anchoring deviation in the Social Anchor condition.*

4 Discussion and Conclusion

In this paper, we have presented an incentivised online experiment that studies the prevalence and persistence of non-numerical and social anchoring effects in a rating environment. Using a framed rating task, we isolated the post-purchase and post-consumption provision of ratings. We employ a double control design to further isolate framing and anchoring effects by using an unframed task as the first benchmark, and a framed, but unanchored, rating task as the second benchmark. In our study's anchoring conditions, we feature high, low, and socially derived anchors presented as visual cues.

We observe significant overrating in all framed rating task conditions compared to the unframed task which induces (expected) payoff maximising decisions. This is consistent with other studies that scrutinise rating behaviour under uncertainty. Buyers tend to provide lenient ratings, which in turn can exacerbate the upward compression of ratings (Bolton et al., 2019). This is also a more general pattern in rating behaviour that has been uncovered previously, with usually a large fraction of positive ratings and very few negative ratings (Hu et al., 2009; Lafky, 2014).

By comparing conditions featuring the rating task, we uncover significant anchoring effects. We observe rating inflation compared with the unanchored rating task for high and social anchors. The effect of the high anchors is most present in early rounds and declining over the course of the experiment, while the effect of the social anchors is concentrated on rounds in which a high social anchor is present. However, the anchoring effect is asymmetric, as we find no effect for low anchors or, respectively, low social anchors. This is comparable with Thorsteinson et al. (2008) who find a stronger effect for high anchors compared to low anchors in performance judgments. However, our findings are in contrast to Jung et al. (2016) who report a stronger effect for low anchors compared to high anchors in "pay what you want" pricing. We conclude that asymmetric anchoring effects might be task specific and may reinforce existing tendencies, such as overrating in quality ratings and performance judgments and the willingness to pay little in "pay what you want" schemes.[11]

The socially derived anchor takes on varying roles between high and low anchors and is indeed directly predictive of the observed ratings. Like De Wilde et al. (2018), we find that social anchors are perceived as more relevant even though they do not contain any informational value in our experiment. We can think of two potential explanations for this observation. First, each participant contributes to the social anchor, potentially causing them to overweigh their own rating. This, in turn, may increase the perceived importance and trust in the social anchor. Second, the social anchor might create an illusion of the

[11] Jung et al. (2016) discuss that low anchors serve as a "license" for low payments to overcome social norms that deem low payments as stingy.

wisdom of crowds inducing more trust in the endogenously derived anchor. Whether either of these two explanations hold, is subject to future research.

Our study indicates that anchoring is a potential channel contributing to the positive skewness of ratings. Whereas it is seemingly possible to inflate the ratings by means of high anchors, low anchors have no significant effect on ratings. Biased ratings could hamper the informativeness of ratings and send wrong signals and diffuse inaccurate information about product quality and service quality resulting in erroneous decisions. This might result in potential welfare losses for consumers, as they purchase products they do not need, pay too much, or buy products of inferior quality due to inflated ratings (Bolton et al., 2019), and for firms and online market platforms, as they could suffer reputational damage when consumers' expectations are not met (Nosko and Tadelis, 2015). Since many online market platforms rely on truthful ratings as part of their marketing strategy to create and promote an accurately reflected reputation (Chen and Xie, 2008), it is important to understand how to promote truthful ratings. Our results highlight that online sellers and market platforms should carefully design the user interface of their rating systems and might also consider including other factors, such as the number of explicit complaints and returns when evaluating the quality of their offers.

Although trust in social anchors is significantly higher on average than trust in the high and low anchors, our design allows us to examine trust only as a moderator. Future research could build on this by systematically investigating trust as a potential mediating mechanism.

While we focused on the fundamental phenomenon of non-numerical anchoring, notably, many rating environments in the real-world feature a hybrid representation, e.g., a 5-star rating system. A promising future avenue for research could be to focus on non-numerical but countable rating systems in the field or field-like environments. Our present study suggests that in such settings visual anchoring contributes to systematic overrating. At the same time, the effects we identify in our abstract and tightly controlled environment are relatively small, implying that detecting corresponding effects in natural settings would likely require substantial data. Large-scale field experiments or natural experiments, e.g., prompted by changes in user-interface design or platform policies, may be best suited to identify and validate anchoring phenomena in real market environments. Our findings may also have broader implications for other online judgment tasks, including survey responses or risk assessments for investment portfolios.

References

Alevy, J. E., Landry, C. E., and List, J. A. (2015). Field experiments on the anchoring of economic valuations. Economic Inquiry, 53(3):1522–1538.

Ariely, D., Loewenstein, G., and Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. The Quarterly Journal of Economics, 118(1):73–106.

Bischoff, I. and Egbert, H. (2013). Social information and bandwagon behavior in voting: An economic experiment. Journal of Economic Psychology, 34:270–284.

Bohnet, I. and Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. American Economic Review, 94(2):362–366.

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How effective are electronic reputation mechanisms? an experimental investigation. Management Science, 50(11):1587–1602.

Bolton, G. E., Kusterer, D. J., and Mans, J. (2019). Inflated reputations: uncertainty, leniency, and moral wiggle room in trader feedback systems. Management Science, 65(11):5371–5391.

Burlig, F., Preonas, L., and Woerman, M. (2020). Panel data and experimental design. Journal of Development Economics, 144:102458.

Chapman, G. B. and Johnson, E. J. (1999). Anchoring, activation, and the construction of values. Organizational Behavior and Human Decision Processes, 79(2):115–153.

Cheah, B. C. (2009). Clustering standard errors or modeling multilevel data. University of Columbia, pages 2–4.

Chen, D. L., Schonger, M., and Wickens, C. (2016). Otree—an open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9:88–97.

Chen, Y. and Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management Science, 54(3):477–491.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research, 43(3):345–354.

Chui, P. M., Fong, L. H. N., Ren, J., and Tam, L. H. (2022). Anchoring effects in repeated auctions of homogeneous objects: Evidence from macao. Journal of Economic Psychology, 90:102514.

Coker, B. L. (2012). Seeking the opinions of others online: Evidence of evaluation overshoot. Journal of Economic Psychology, 33(6):1033–1042.

De Wilde, T. R., Ten Velden, F. S., and De Dreu, C. K. (2018). The anchoring-bias in groups. Journal of Experimental Social Psychology, 76:116–126.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science, 49(10):1407–1424.

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive Marketing, 21(4):23–45.

Epley, N. and 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.

Fisher, M., Newman, G. E., and Dhar, R. (2018). Seeing stars: How the binary bias distorts the interpretation of customer ratings. Journal of Consumer Research, 45(3):471–489.

Frederick, S. W. and Mochon, D. (2012). A scale distortion theory of anchoring. Journal of Experimental Psychology: General, 141(1):124.

Fudenberg, D., Levine, D. K., and Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. American Economic Journal: Microeconomics, 4(2):131–45.

Furnham, A. and Boo, H. C. (2011). A literature review of the anchoring effect. The Journal of Socio-economics, 40(1):35–42.

Givi, J., Grossman, D. M., and Kardes, F. R. (2024). Erroneous consumer evaluations in the marketplace: How consumer evaluations are biased by raw performance scores. Psychology & Marketing, 41(3):665–676.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. Journal of the Economic Science Association, 1(1):114–125.

Hoffart, J. C., Olschewski, S., and Rieskamp, J. (2019). Reaching for the star ratings: A bayesian-inspired account of how people use consumer ratings. Journal of Economic Psychology, 72:99–116.

Hox, J., Moerbeek, M., and Van de Schoot, R. (2017). Multilevel analysis: Techniques and applications. Routledge.

Hu, N., Zhang, J., and Pavlou, P. A. (2009). Overcoming the j-shaped distribution of product reviews. Communications of the ACM, 52(10):144–147.

Ivanova-Stenzel, R. and Seres, G. (2021). Are strategies anchored? European Economic Review, 135:103725.

Ivanova-Stenzel, R. and Seres, G. (2022). Anchored strategic reasoning. Economics Letters, 212:110330.

Jacowitz, K. E. and Kahneman, D. (1995). Measures of anchoring in estimation tasks. Personality and Social Psychology Bulletin, 21(11):1161–1166.

Jung, M. H., Perfecto, H., and Nelson, L. D. (2016). Anchoring in payment: Evaluating a judgmental heuristic in field experimental settings. Journal of Marketing Research, 53(3):354–368.

Kimbrough, E. O., Porter, D., and Schneider, M. (2021). Reference dependent prices in bargaining: An experimental examination of precise first offers. Journal of Economic Psychology, 86:102406.

Lafky, J. (2014). Why do people rate? Theory and evidence on online ratings. Games and Economic Behavior, 87:554–570.

LeBoeuf, R. A. and Shafir, E. (2006). The long and short of it: Physical anchoring effects. Journal of Behavioral Decision Making, 19(4):393–406.

Li, L., Tadelis, S., and Zhou, X. (2020). Buying reputation as a signal of quality: Evidence from an online marketplace. The RAND Journal of Economics, 51(4):965–988.

Magnani, M. (2020). The economic and behavioral consequences of online user reviews. Journal of Economic Surveys, 34(2):263–292.

Maniadis, Z., Tufano, F., and List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. American Economic Review, 104(1):277–90.

McHugh, S. and Ranyard, R. (2016). Consumers' credit card repayment decisions: The role of higher anchors and future repayment concern. Journal of Economic Psychology, 52:102–114.

Meub, L. and Proeger, T. (2016). Can anchoring explain biased forecasts? experimental evidence. Journal of Behavioral and Experimental Finance, 12:1–13.

Meub, L. and Proeger, T. E. (2015). Anchoring in social context. Journal of Behavioral and Experimental Economics, 55:29–39.

Moe, W. W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. Journal of Marketing Research, 48(3):444–456.

Northcraft, G. B. and Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. Organizational Behavior and Human Decision Processes, 39(1):84–97.

Nosko, C. and Tadelis, S. (2015). The limits of reputation in platform markets: An empirical analysis and field experiment (no. w20830). Technical report, National Bureau of Economic Research.

Oppenheimer, D. M., LeBoeuf, R. A., and Brewer, N. T. (2008). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. Cognition, 106(1):13–26.

Phillips, O. R. and Menkhaus, D. J. (2010). The culture of private negotiation: Endogenous price anchors in simple bilateral bargaining experiments. Journal of Economic Behavior & Organization, 76(3):705–715.

Slovic, P. and Lichtenstein, S. (1972). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6(6):649–744.

Sugden, R., Zheng, J., and Zizzo, D. J. (2013). Not all anchors are created equal. Journal of Economic Psychology, 39:21–31.

Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., and Privette, M. (2008). Anchoring effects on performance judgments. Organizational Behavior and Human Decision Processes, 107(1):29–40.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157):1124–1131.

Ünveren, B. and Baycar, K. (2019). Historical evidence for anchoring bias: The 1875 cadastral survey in istanbul. Journal of Economic Psychology, 73:1–14.

Yoon, S. and Fong, N. (2019). Uninformative anchors have persistent effects on valuation judgments. Journal of Consumer Psychology, 29(3):391–410.

Yoon, S., Fong, N. M., and Dimoka, A. (2019). The robustness of anchoring effects on preferential judgments. Judgment & Decision Making, 14(4).

5 Appendix

5.1 Derivation of optimal rating r^*

A rational decision maker wants to maximize her expected utility $E(U) = A - E(|r - q|)$, where A is the maximum attainable utility, when the rating r coincides with the uniformly distributed quality $q \sim \mathcal{U}(q_L, q_H)$, where q_L is the lower bound of the quality interval and q_H is the upper bound of the quality interval. Given the linear payment rule, the absolute difference between the rating and quality $|r - q|$ is subtracted from A. The decision problem is

$$\max_r E(A - |r - q|).$$

We consider three cases *i)* $r < q_L$, *ii)* $r > q_H$ and *iii)* $r \in [q_L, q_H]$. In case *i)* it is immediate that any $r < q_L$ is strictly dominated by setting $r' = q_L$, as the utility is larger by $q_L - r$ for any possible q. Similarly in case *ii)* it is immediate that $r > q_H$ is strictly dominated by $r'' = q_H$. The optimal rating r^* must therefore be within the interval $[q_L, q_H]$. We rewrite the expected utility accordingly:

$$E(A - |r - q|) = A - \frac{1}{q_H - q_L} \int_{q_L}^{r} r - \tilde{q} \; d\tilde{q} - \frac{1}{q_H - q_L} \int_{r}^{q_H} \tilde{q} - r \; d\tilde{q},$$

where solving the integrals yields

$$E(A - |r - q|) = A - \frac{1}{q_H - q_L}(r^2 - rq_L - rq_H + \frac{q_L^2 + q_H^2}{2}).$$

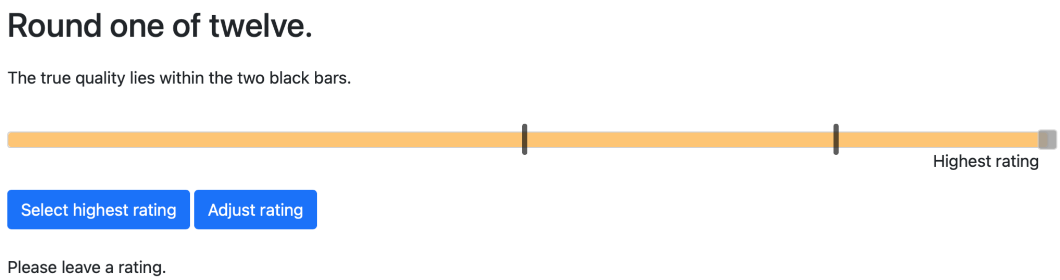
We solve the first-order condition with respect to r to find r^*:

$$\frac{\partial E(A - |r - q|)}{\partial r} = -\frac{2r - q_L - q_H}{q_H - q_L} \stackrel{!}{=} 0 \iff r^* = \frac{q_H + q_L}{2}.$$

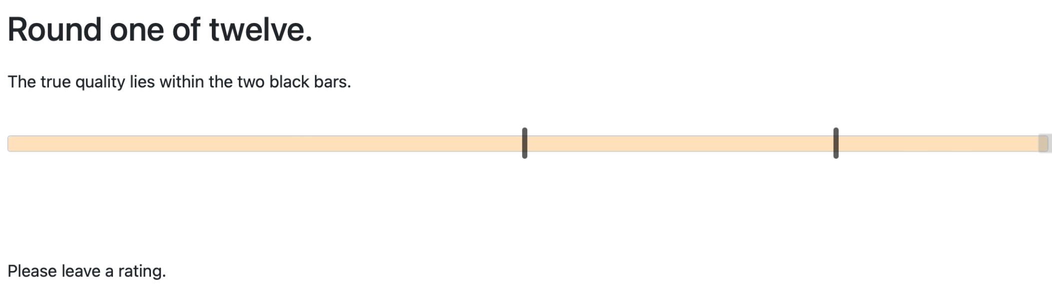
The second-order condition verifies that r^* is indeed a local maximiser since

$$\frac{\partial^2 E(A - |r - q|)}{\partial r^2} = -2 < 0.$$

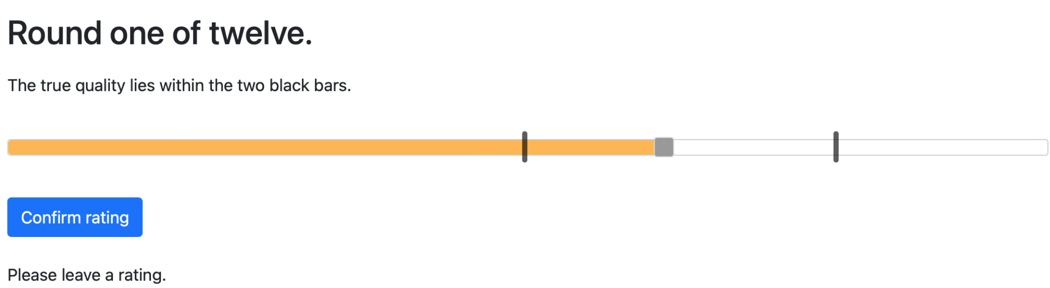
5.2 Decision screen



(a) Initial decision screen.



(b) After clicking on “Adjust rating”.



(c) After initiating a new rating.

Figure 5: Decision screen in the rating task (High Anchor condition).

5.3 Post-experimental questionnaire

Cognitive reflection test

1. A bat and a ball cost 1.10 euros. The bat costs one euro more than the ball. How many euro cents does the ball cost? [ ]

2. 5 machines need 5 minutes to produce 5 products. How many minutes does it take 100 machines to produce 100 products? [ ]

3. The water lilies in a pond double in size every day. If after 48 days the lake is completely covered with water lilies, how many days did it take until it was half covered? [ ]

Statistical aptitude questions

1. Do you have prior knowledge of statistics? [Yes / No]

2. Do you know what an expected value is? [Yes / No]

3. Consider a fair 6-sided die. If the number rolled is a 5 or 6, you win 6 euros. If the number rolled is 4 or less, you win 3 euros. Please determine the expected profit for rolling the die once. [ ]

Perceived relevance of anchors (Only in anchoring conditions)

1. The displayed rating in each round helped me with my evaluation.

   [I do not agree ○ ○ ○ ○ ○ ○ ○ I agree completely]

2. The displayed rating in each round was informative for me.

   [I do not agree ○ ○ ○ ○ ○ ○ ○ I agree completely]

3. I based my decision on the displayed rating and this influenced my decision.

   [I do not agree ○ ○ ○ ○ ○ ○ ○ I agree completely]

5.4 Robustness checks

5.4.1 No exclusion criteria

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
High Anchor	0.087***	0.119***	0.058*	0.095**
	(0.018)	(0.025)	(0.023)	(0.029)
Low Anchor	-0.015	-0.062†	-0.044*	-0.086*
	(0.019)	(0.033)	(0.023)	(0.036)
Social Anchor	0.074***	0.082***	0.045*	0.058*
	(0.016)	(0.022)	(0.020)	(0.026)
No Anchor	0.030*	0.025		
	(0.013)	(0.015)		
Second half		-0.003		0.007
		(0.004)		(0.019)
High Anchor × Second half		-0.064*		-0.074*
		(0.029)		(0.035)
Low Anchor × Second half		0.094*		0.084*
		(0.037)		(0.041)
Social Anchor × Second half		-0.016		-0.026
		(0.027)		(0.033)
No Anchor × Second half		0.010		
		(0.019)		
Constant	0.015†	0.016†	0.047**	0.043*
	(0.008)	(0.009)	(0.016)	(0.017)
Intraclass correlation	0.030***	0.031***	0.029***	0.030***
Observations	2952	2952	2364	2364
Number of subjects	246	246	197	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by multi-level model with subject-level random-effects. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.023***	0.984***	1.034***	1.034***
	(0.0138)	(0.0161)	(0.0187)	(0.0156)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.113***	0.0229	-0.117***	-0.0266*
	(0.0192)	(0.0357)	(0.0342)	(0.0124)
χ^2-statistic ($\alpha = 1$)	2.86	1.02	3.26	4.65
Prob. > χ^2	0.091	0.311	0.071	0.031
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	550	225	325	588
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by multi-level model with subject-level random-slopes. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
Marked Rounds	-0.003	-0.003	-0.002	-0.002
	(0.004)	(0.004)	(0.023)	(0.023)
Social Anchor	-0.014	-0.014	-0.045†	-0.043†
	(0.021)	(0.021)	(0.025)	(0.025)
High Anchor		0.088***		0.060*
		(0.025)		(0.029)
Low Anchor		-0.043		-0.072*
		(0.029)		(0.032)
No Anchor		0.029*		
		(0.014)		
Social Anchor × Marked Rounds	0.177***	0.177***	0.176***	0.176***
	(0.033)	(0.033)	(0.040)	(0.040)
High Anchor × Marked Rounds		-0.002		-0.003
		(0.032)		(0.039)
Low Anchor × Marked Rounds		0.057		0.056
		(0.041)		(0.047)
No Anchor × Marked Rounds		0.001		
		(0.023)		
Constant	0.014	0.016†	0.052**	0.048**
	(0.009)	(0.008)	(0.018)	(0.017)
Intraclass correlation	0.040***	0.032***	0.039***	0.032***
Observations	1188	2952	1188	2364
Number of subjects	99	246	99	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by multi-level model with subject-level random-effects. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. * *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.022***	0.984***	1.029***	1.034***
	(0.0138)	(0.0160)	(0.0201)	(0.0156)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.0562*	0.0434	-0.0361	-0.0252†
	(0.0285)	(0.0457)	(0.0424)	(0.0134)
Deviation × trust $(r_t^* - \bar{r}_{t-1})s_i$	-0.150*	-0.0559	-0.205*	-0.00802
	(0.0626)	(0.0649)	(0.0933)	(0.0196)
χ^2-statistic ($\alpha = 1$)	2.67	0.95	2.13	4.68
Prob. > χ^2	0.103	0.331	0.145	0.031
χ^2-statistic ($\beta + \gamma = 0$)	21.89***	0.06	11.87***	3.19†
Prob. > χ^2	0.0000	0.8062	0.0006	0.0740
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	550	225	325	588
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by multilevel model with subject-level random-slopes. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

5.4.2 OLS regressions

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
High Anchor	0.057***	0.073***	0.034*	0.055*
	(0.013)	(0.017)	(0.017)	(0.022)
Low Anchor	0.047***	0.034*	0.023	0.015
	(0.010)	(0.016)	(0.015)	(0.020)
Social Anchor	0.067***	0.075***	0.043*	0.056*
	(0.018)	(0.021)	(0.021)	(0.025)
No Anchor	0.023*	0.019		
	(0.011)	(0.013)		
Second half		0.001		0.010
		(0.001)		(0.014)
High Anchor × Second half		-0.032†		-0.041†
		(0.018)		(0.023)
Low Anchor × Second half		0.025		0.016
		(0.020)		(0.025)
Social Anchor × Second half		-0.017		-0.026
		(0.026)		(0.029)
No Anchor × Second half		0.009		
		(0.014)		
Constant	0.008	0.008	0.033**	0.028†
	(0.006)	(0.007)	(0.012)	(0.015)
Adj. R^2	0.011	0.011	0.003	0.004
Observations	2829	2829	2242	2242
Number of subjects	246	246	197	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by OLS regression with clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.030***	1.007***	1.033***	1.009***
	(0.0154)	(0.0203)	(0.0237)	(0.0113)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.0788***	-0.0178	-0.0965**	-0.0259*
	(0.0173)	(0.0432)	(0.0331)	(0.0113)
F-statistic ($\alpha = 1$)	3.89	0.11	1.99	0.59
Prob. > F	0.054	0.739	0.165	0.447
Adj. R^2	0.976	0.987	0.958	0.987
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	462	193	269	576
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by OLS regression with suppressed constant term. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

	Base category: No Framing		Base category: No Anchor	
	(1)	(2)	(3)	(4)
Marked Rounds	-0.006***	-0.006***	0.001	0.001
	(0.001)	(0.001)	(0.017)	(0.017)
Social Anchor	0.007	0.008	-0.013	-0.012
	(0.022)	(0.022)	(0.025)	(0.025)
High Anchor		0.044**		0.025
		(0.015)		(0.019)
Low Anchor		0.029†		0.009
		(0.016)		(0.019)
No Anchor		0.020†		
		(0.011)		
Social Anchor × Marked Rounds	0.118***	0.119***	0.111**	0.111**
	(0.031)	(0.031)	(0.036)	(0.036)
High Anchor × Marked Rounds		0.025		0.018
		(0.022)		(0.028)
Low Anchor × Marked Rounds		0.036		0.029
		(0.026)		(0.032)
No Anchor × Marked Rounds		0.007		
		(0.017)		
Constant	0.016†	0.011†	0.038*	0.033*
	(0.009)	(0.007)	(0.015)	(0.013)
Adj. R^2	0.062	0.022	0.027	0.015
Observations	1099	2829	1097	2242
Number of subjects	99	246	99	197

The dependent variable is the normalised rating $\tilde{r}_{it}$. Estimation by OLS regression with clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. * *, ** and *** denote significance at the 5%, 1% and 0.1% level, respectively.

	Social Anchor			High Anchor
	(1)	(2)	(3)	(4)
Optimal rating r_t^*	1.030***	1.008***	1.031***	1.009***
	(0.0155)	(0.0201)	(0.0244)	(0.0113)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.0380	0.00365	-0.0378	-0.0227†
	(0.0264)	(0.0526)	(0.0434)	(0.0114)
Deviation × trust $(r_t^* - \bar{r}_{t-1})s_i$	-0.111†	-0.0588	-0.151	-0.0179
	(0.0604)	(0.0644)	(0.0975)	(0.0172)
F-statistic $(\alpha = 1)$	3.78	0.15	1.66	0.62
Prob. $> F$	0.058	0.702	0.203	0.437
F-statistic $(\beta + \gamma = 0)$	11.91**	1.01	6.51*	4.81*
Prob. $> F$	0.0012	0.3190	0.0139	0.0331
Adj. R^2	0.977	0.987	0.958	0.987
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$	all
Observations	462	193	269	576
Number of subjects	50	50	50	49

The dependent variable is the submitted rating r_{it}. Estimation by OLS regression with suppressed constant term. Clustered standard errors in parentheses. Controls are the CRT and statistical aptitude scores. † denotes marginal significance at the 10% level. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

Translated instructions and review questions

[Text in brackets was not observed by subjects. Presented sliders were interactive in the digital instructions. For consecutive pictures of sliders only the first was visible to subjects, the remaining are exemplary of the interaction.]

Thank you for your participation in today's experiment. Please do not communicate with other participants during the experiment. Throughout the entire duration of the experiment, please only use the experiment programme that will be displayed to you and please do not use any other programmes or applications on your computer. You can earn money in this experiment. The exact amount depends on your decisions and the other participants decisions. If you have any questions during the experiment, please use the chat function on Zoom to contact one of the experimenters.

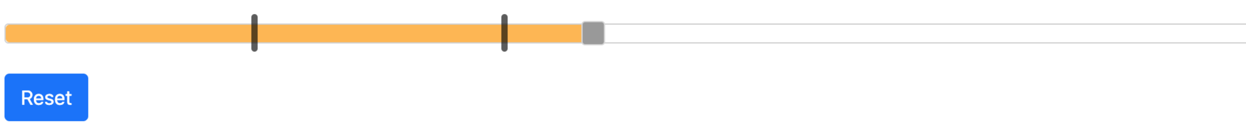
In this experiment you will make simple decisions on your computer. All decisions will remain anonymous. This means, that you will never learn the identity of the other participants and none of the other participants will learn your identity. All monetary values will be displayed in Experimental Currency Units [ECU].

[No Anchor condition]
Your Task:
In every round of this experiment, you will be presented with a quality interval which is located on a bar. The bar represents the entire quality range. The true quality is always contained in this interval and each point within the range is equally likely. The quality is increasing from left to right along the bar.

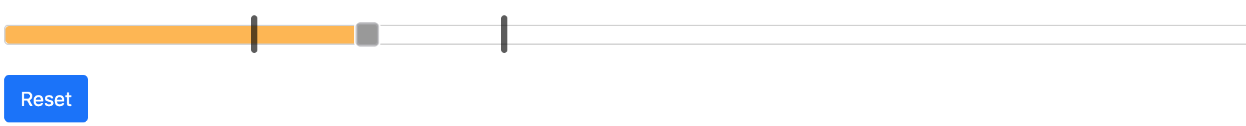
Your task consists of rating the quality you obtain. You can provide a rating with the help of the handle on the bar. You can initialise your rating by clicking on the bar. Then you can move the handle on the bar to the left or right and bring it to the desired position. Below is an example of how to operate the handle.

Please enter a rating

[Examples of interaction]



Please enter a rating: Your rating is **outside** the quality interval.

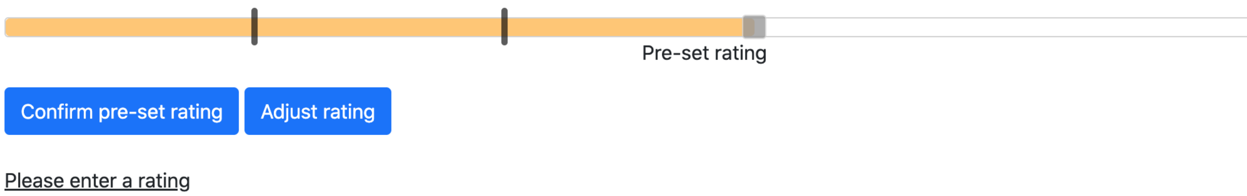


Please enter a rating: Your rating is **inside** the quality interval.

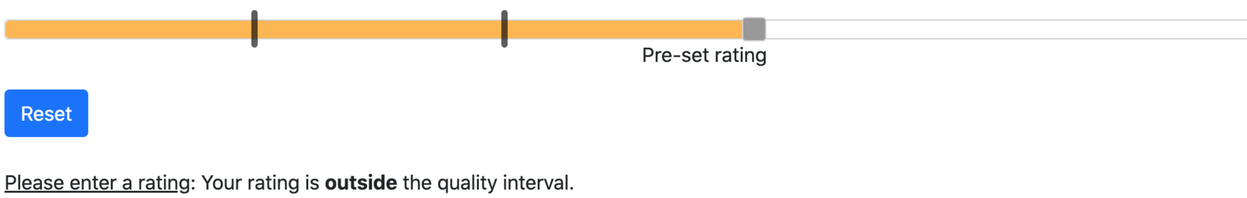
[End examples]

[High Anchor and Low Anchor conditions]
Your Task:
In every round of this experiment, you will be presented with a quality interval which is located on a bar. The bar represents the entire quality range. The true quality is always contained in this interval and each point within the range is equally likely. The quality is increasing from left to right along the bar.

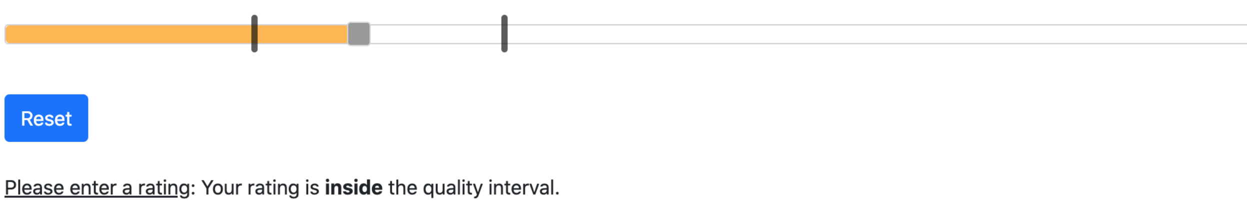
Your task consists of rating the quality you obtain. You can provide a rating with the help of the handle on the bar. First, you can choose whether you want to choose a pre-set rating or adjust the rating. If you choose the pre-set rating in a round, you cannot revoke it. If you want to adjust the rating, you must initialise your rating by clicking on the bar. Then you can move the handle on the bar to the left or right and bring it to the desired position. Below is an example of how to operate the handle.



[Example of interaction after 'confirm pre-set rating']



[Examples of interaction after 'adjust rating']



[End examples]

[Social Anchor condition]
Your Task:
In this experiment, you will be assigned to a group consisting of you and four other participants. This group remains the same and does not change throughout the twelve rounds of the experiment. You and all other party members are shown a quality interval on a bar each round. The true quality lies in this interval, with each value in the interval being equally likely. The quality is increasing from left to right along the bar. The quality interval and true quality are the same for all group members.

Your task consists of rating the quality you obtain. You can provide a rating with the help of the handle on the bar. You can initialise your rating by clicking on the bar. Then you can move the handle on the bar to the left or right and bring it to the desired position. Below is an example of how to operate the handle in the first round.

Please enter a rating

[Example of interaction]

Reset

Please enter a rating: Your rating is **inside** the quality interval.

[End example]

Beginning with the second round, the average rating of your group (including your rating) from the previous round will be displayed next to the quality interval. First, you must decide to either choose the average rating of the previous round or to adjust the rating. If you choose the average rating of the previous round in a round, you cannot revoke it. If you want to adjust the rating, you must initialise your rating by clicking on the bar. Then you can move the handle on the bar to the left or right and bring it to the desired position. Below is an example of how to operate the handle from the second round onwards.

Average rating of the previous round

Choose the average rating of the previous round | Adjust rating

Please enter a rating

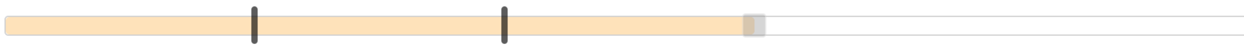
[Example of interaction after 'Choose average rating of the previous round']

Average rating of the previous round

Reset

Please enter a rating: Your rating is **outside** the quality interval.

[Examples of interaction after ‘Adjust rating’]



Please enter a rating

Reset

Please enter a rating: Your rating is **inside** the quality interval.

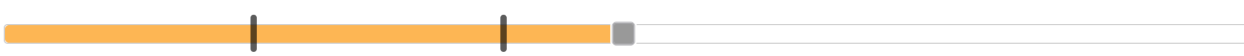
[End examples]

[No Framing condition]
Your Task: In every round of this experiment, you will be presented with an interval which is located on a bar. The target value is exactly in the middle of the interval.

Your task consists of hitting this target value as closely as possible. You can make an entry with the help of the handle on the bar. You can initialise your entry by clicking on the bar. Then you can move the handle on the bar to the left or right and bring it to the desired position. Below is an example of how to operate the handle.

Please make an entry

[Examples of interaction]



Reset

Please make an entry: Your entry is **outside** the interval.

Reset

Please make an entry: Your entry is **inside** the interval.

[End examples]

[End conditions]

[No Anchor, High Anchor, Low Anchor and Social Anchor conditions]

Your Payoff:

There are twelve rounds in total. In each round, the quality interval, and thus the true quality, is independent of the previous round. After completing the twelve rounds, a round will be chosen at random, which will determine your payoff. The amount of money in this round depends on how close you get to true quality, i.e. the closer your rating is to the true quality, the higher are your earnings in the round. Your earnings break down as follows: one thousand ECU less the deviation from the true quality. Depending on the true quality and your rating, your payoff will range from zero ECU to one thousand ECU. In the following, you find an example of your payoff depending on your rating and the true quality.

Please enter a rating, Your Payoff (in ECU) in this case would be:

Please enter a rating, Your Payoff (in ECU) in this case would be: 726

[No Framing condition]

Your Payoff:

There are twelve rounds in total. In each round you see a new interval. After completing the twelve rounds, a round will be chosen at random, which will determine your payoff. The amount of money in this round depends on how close you get to target value, i.e. the closer your entry is to the target value, the higher are your earnings in the round. Your earnings break down as follows: one thousand ECU less the deviation from the target value. Depending on the target value and your entry, your payoff will range from zero ECU to one thousand ECU. In the following, you find an example of your payoff depending on your entry and the target value.

Please enter a rating, Your payoiff (in ECU) in this case would be:

Please enter a rating, Your payoiff (in ECU) in this case would be: 832

[End conditions]

The ECU collected during the experiment will be paid out in euros after the experiment. One hundred ECU equals one euro. For taking part in today's experiment you will also receive a participation fee of two euros. Your earnings from this experiment will be paid to you via PayPal no later than the day after the experiment.

[No Anchor, High Anchor, Low Anchor and Social Anchor conditions]
Review Questions:
1. Which of the following statements is true?

   a.) The true quality is always contained in the quality interval within the two black bars.

   b.) The true quality is not always contained in the quality interval within the two black bars.

2. Which of the following statements is true?

   a.) The further the rating is from the true quality, the higher is my payoff.

   b.) The closer the rating is to the true quality, the higher is my payoff.

3. Which of the following statements is true?

   a.) The quality intervals and the true quality are independent between rounds.

   b.) The quality intervals and the true quality are interdependent between rounds

[No Framing condition]
Review Questions:
1. Which of the following statements is true?

   a.) The target value is always contained in the quality interval within the two black bars.

   b.) The target value is not always contained in the quality interval within the two black bars.

2. Which of the following statements is true?

   a.) The further the entry is from the target value, the higher is my payoff.

   b.) The closer the entry is to the target value, the higher is my payoff.

3. Which of the following statements is true?

   a.) The target value is always exactly in the middle of the interval.

   b.) The target value takes on a random value within the interval.

[End experimental conditions]

Original instructions and review questions

[Text in brackets was not observed by subjects. Presented sliders were interactive in the digital instructions. For examples of the interactions see translated instructions.]

Vielen Dank für Ihre Teilnahme am heutigen Experiment. Während des Experimentes ist es Ihnen nicht erlaubt, mit anderen teilnehmenden Personen zu kommunizieren. Bitte benutzen Sie nur die für das Experiment vorgesehenen Programme und Funktionen und benutzen Sie während des Experimentes keine weiteren Anwendungen auf Ihrem Computer. Außerdem können Sie mit den Aktionen, die Sie während des Experiments durchführen, Geld verdienen. Der genaue Betrag, den Sie erhalten, wird während des Experimentes festgelegt und hängt von Ihren Entscheidungen und den Entscheidungen anderer ab. Wenn Sie während des Experiments Fragen haben, melden Sie sich bitte über die Chatfunktion bei Zoom und warten Sie, bis die/der Experimentator:in sich bei Ihnen meldet.

In diesem Experiment werden Sie einfache Entscheidungen am Computer treffen. Alle Entscheidungen bleiben anonym. Das heißt, Sie erfahren die Identität der anderen Teilnehmer nicht und kein Teilnehmer erfährt Ihre Identität. Sämtliche Geldangaben innerhalb des Experiments werden in ECU (Experimental Currency Unit) angegeben.

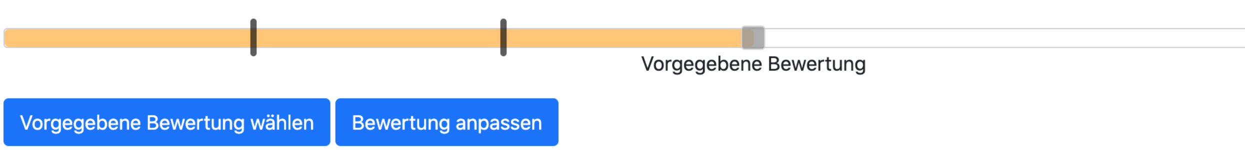
[No Anchor condition]
Ihre Aufgabe:
In diesem Experiment wird Ihnen jede Runde ein Qualitätsintervall angezeigt, welches auf einem Balken liegt. Die wahre Qualität liegt in diesem Intervall, wobei jeder Wert im Intervall gleichwahrscheinlich ist. Die Qualität steigt auf dem Balken von links nach rechts an.

Ihre Aufgabe besteht darin, diese Qualität, die Sie erhalten, zu bewerten. Die Bewertung nehmen Sie mit Hilfe eines Reglers auf dem Balken vor. Durch einen Mausklick auf den Balken initialisieren Sie ihre Bewertung. Sie können anschließend den Regler weiter bewegen. Im Folgenden finden Sie ein Beispiel zur Bedienung des Reglers.

Bitte geben Sie eine Bewertung ab

[High Anchor and Low Anchor conditions]
Ihre Aufgabe:
In diesem Experiment wird Ihnen jede Runde ein Qualitätsintervall angezeigt, welches auf einem Balken liegt. Die wahre Qualität liegt in diesem Intervall, wobei jeder Wert im Intervall gleichwahrscheinlich ist. Die Qualität steigt auf dem Balken von links nach rechts an.

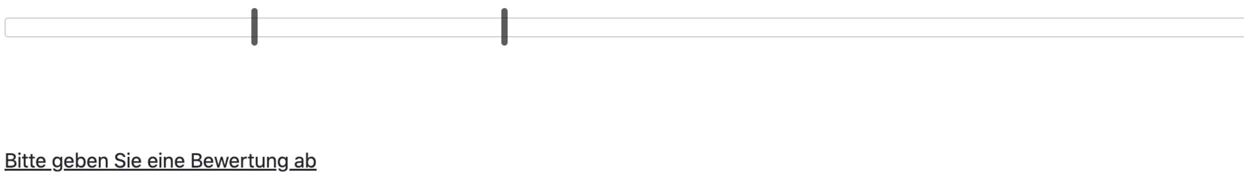
Ihre Aufgabe besteht darin, diese Qualität, die Sie erhalten, zu bewerten. Die Bewertung nehmen Sie mit Hilfe eines Reglers auf dem Balken vor. Zunächst können Sie wählen, ob sie eine vorgegebene Bewertung wählen oder die Bewertung anpassen möchten. Sofern Sie in einer Runde die vorgegebene Bewertung wählen, können Sie diese nicht widerrufen. Sofern Sie die Bewertung anpassen möchten, müssen Sie durch einen Mausklick auf den Balken Ihre Bewertung initialisieren. Sie können anschließend den Regler weiter bewegen. Im Folgenden finden Sie ein Beispiel zur Bedienung des Reglers.



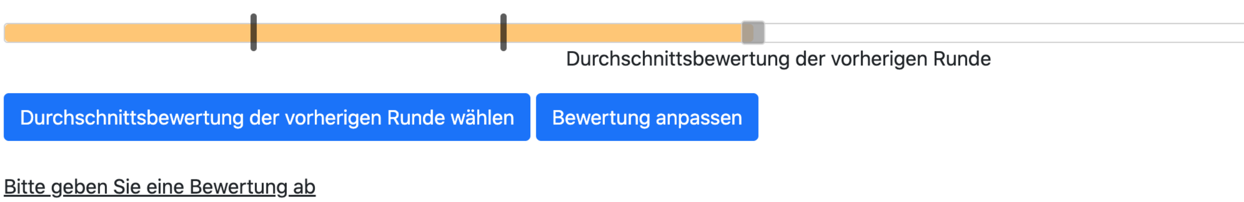
Bitte geben Sie eine Bewertung ab

[Social Anchor condition]
Ihre Aufgabe:
In diesem Experiment werden sie einer Gruppe zugeordnet, welche aus Ihnen und vier weiteren Teilnehmern besteht. Diese Gruppe bleibt während der zwölf Runden des Experiments gleich und ändert sich nicht. Ihnen und allen anderen Gruppenmitgliedern wird jede Runde ein Qualitätsintervall auf einem Balken angezeigt. Die wahre Qualität liegt in diesem Intervall, wobei jeder Wert im Intervall gleichwahrscheinlich ist. Die Qualität steigt auf dem Balken von links nach rechts an. Das Qualitätsintervall und die wahre Qualität sind für alle Gruppenmitglieder identisch.

Ihre Aufgabe besteht darin, diese Qualität, die Sie erhalten, zu bewerten. Die Bewertung nehmen Sie mit Hilfe eines Reglers auf dem Balken vor. Durch einen Mausklick auf den Balken initialisieren Sie Ihre Bewertung. Sie können anschließend den Regler weiter bewegen. Im Folgenden finden Sie ein Beispiel zur Bedienung des Reglers in der ersten Runde.

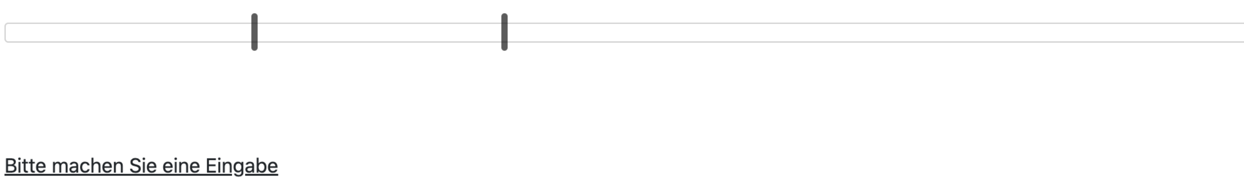


Ab der zweiten Runde wird Ihnen neben dem Qualitätsintervall die Durchschnittsbewertung Ihrer Gruppe (inklusive Ihrer Bewertung) aus der vorherigen Runde angezeigt. Sie können zunächst wählen, ob sie die Durchschnittsbewertung der vorherigen Runde wählen oder die Bewertung anpassen möchten. Sofern Sie in einer Runde die Durchschnittsbewertung der vorherigen Runde wählen, können Sie diese nicht widerrufen. Sofern Sie die Bewertung anpassen möchten, müssen Sie durch einen Mausklick auf den Balken Ihre Bewertung initialisieren. Sie können anschließend den Regler weiter bewegen. Im Folgenden finden Sie ein Beispiel zur Bedienung des Reglers ab der zweiten Runde.



[No Framing condition]
Ihre Aufgabe: In diesem Experiment wird Ihnen jede Runde ein Intervall angezeigt, welches auf einem Balken liegt. Genau in der Mitte innerhalb des Intervalls ist der Zielwert.

Ihre Aufgabe besteht darin, diesen Zielwert mit einer Eingabe auf dem Balken so genau wie möglich zu treffen. Die Eingabe nehmen Sie mit Hilfe eines Reglers auf dem Balken vor. Durch einen Mausklick auf den Balken initialisieren Sie ihre Eingabe. Sie können anschließend den Regler weiter bewegen. Im Folgenden finden Sie ein Beispiel zur Bedienung des Reglers.



[End experimental conditions]

[No Anchor, High Anchor, Low Anchor and Social Anchor conditions]
Ihre Auszahlung:

Insgesamt gibt es zwölf Runden. In jeder Runde ist das Qualitätsintervall und damit die wahre Qualität unabhängig von der vorherigen Runde. Nach Abschluss der zwölf Runden wird zufällig eine Runde gewählt, welche Ihre Auszahlung bestimmt. Der Geldbetrag dieser Runde hängt davon ab wie nahe Sie an die wahre Qualität kommen, d.h. je näher Ihre Bewertung an der wahren Qualität ist, desto höher ist Ihr Verdienst in der Runde. Ihr Verdienst setzt sich wie folgt zusammen: eintausend ECU abzüglich der Abweichung von der wahren Qualität. Abhängig von der wahren Qualität und Ihrer Bewertung, liegt Ihre Auszahlung zwischen null ECU und eintausend ECU. Im Folgenden wird Ihnen Ihre Auszahlung in Abhängigkeit ihrer Bewertung und der wahren Qualität an einem Beispiel verdeutlicht.

Bitte geben Sie eine Bewertung ab, Ihre Auszahlung (in ECU) wäre in diesem Fall:

[No Framing condition]

Ihre Auszahlung:

Insgesamt gibt es zwölf Runden. In jeder Runde sehen Sie ein neues Intervall. Nach Abschluss der zwölf Runden wird zufällig eine Runde gewählt, welche Ihre Auszahlung bestimmt. Der Geldbetrag dieser Runde hängt davon ab wie nahe Sie mit Ihrer Eingabe an den Zielwert kommen, d.h. je näher Ihre Eingabe am Zielwert ist, desto höher ist Ihr Verdienst in der Runde. Ihr Verdienst setzt sich wie folgt zusammen: eintausend ECU abzüglich der Abweichung vom Zielwert. Abhängig vom Zielwert und Ihrer Eingabe, liegt Ihre Auszahlung zwischen null ECU und eintausend ECU. Im Folgenden wird Ihnen Ihre Auszahlung in Abhängigkeit ihrer Eingabe und dem Zielwert an einem Beispiel verdeutlicht.

Bitte machen Sie eine Eingabe, Ihre Auszahlung (in ECU) wäre in diesem Fall:

[End experimental conditions]

Die während des Experiments gesammelten ECU werden werden im Anschluss an das Experiment in Euro ausgezahlt. Dabei entsprechen zweihundert ECU = einem Euro. Für die Teilnahme am heutigen Experiment erhalten Sie zusätzlich eine Teilnahmevergütung von drei Euro. Ihr Verdienst in diesem Experiment wird Ihnen spätestens am Folgetag des Experiments über PayPal ausgezahlt.

[No Anchor, High Anchor, Low Anchor and Social Anchor conditions]
Kontrollfragen:
1. Welche der folgenden Aussagen trifft zu?

   a.) Die wahre Qualität liegt immer im Qualitätsintervall zwischen den beiden schwarzen Balken.

   b.) Die wahre Qualität liegt nicht immer im Qualitätsintervall zwischen den beiden schwarzen Balken.

2. Welche der folgenden Aussagen trifft zu?

   a.) Je weiter entfernt die Bewertung von der wahren Qualität ist, desto größer ist mein Verdienst.

   b.) Je näher die Bewertung an der wahren Qualität ist, desto größer ist mein Verdienst.

3. Welche der folgenden Aussagen trifft zu?

   a.) Die Qualitätsintervalle und die wahre Qualität sind unabhängig voneinander zwischen den Runden.

   b.) Die Qualitätsintervalle und die wahre Qualität sind abhängig voneinander zwischen den Runden.

[No Framing condition]
Kontrollfragen:
1. Welche der folgenden Aussagen trifft zu?

   a.) Der Zielwert liegt immer im Intervall zwischen den beiden schwarzen Balken.

   b.) Der Zielwert liegt nicht immer im Intervall zwischen den beiden schwarzen Balken.

2. Welche der folgenden Aussagen trifft zu?

   a.) Je weiter entfernt die Eingabe von dem Zielwert ist, desto größer ist mein Verdienst.

   b.) Je näher die Eingabe am Zielwert ist, desto größer ist mein Verdienst.

3. Welche der folgenden Aussagen trifft zu?

   a.) Der Zielwert liegt immer genau in der Mitte des Intervalls.

   b.) Der Zielwert nimmt einen zufälligen Wert innerhalb des Intervalls an.

[End experimental conditions]