UNIVERSITY OF LIVERPOOL | Management School

Working Paper in Economics

202319

Non-numerical and social anchoring in consumer-generated ratings

Yigit Oezcelik*, Michel Tolksdorf†

* Corresponding author. University of Liverpool, University of Liverpool Management School, Mulberry Court, Mulberry Street, Liverpool L69 7ZH, United Kingdom, Yigit.Oezcelik@liverpool.ac.uk. This project received ethics approval by the University of Liverpool Ethics Committee (Project ID 09813).
† Berlin School of Economics and Technische Universität Berlin, H91, Str. des 17. Juni 135, 10623 Berlin, Germany, michel.tolksdorf@tuberlin.de.

Non-numerical and social anchoring in consumer-generated ratings

Yigit Oezcelik[*], Michel Tolksdorf[†‡]

December, 2023

Abstract

Inaccurate online ratings can be harmful to both consumers and firms. We perform an experiment to assess the effect of the anchoring bias on consumer ratings. Our rating task is a framed variation of the slider task. We diverge from the literature by implementing non-numerical (visual) anchors. We compare three anchoring conditions, with either high, low, or socially derived anchors present, against two control conditions – one without anchors and an unframed slider task. High and socially derived anchors lead to significant overrating compared to both control conditions. We find no difference between low anchors and the control condition without anchors, whereas both exhibit overrating compared with the unframed slider task. Participants place higher trust in socially derived anchors compared with high and low anchors. When there is a social context, the trust participants exhibit towards the socially derived anchors explains the anchoring bias.

Keywords: Anchoring, online ratings, laboratory experiment

JEL Codes: C91, D80, D91 **PsycINFO Classification Codes:** 3920, 3940

[*]Corresponding author. University of Liverpool, University of Liverpool Management School, Mulberry Court, Mulberry Street, Liverpool L69 7ZH, United Kingdom, Yigit.Oezcelik@liverpool.ac.uk. This project received ethics approval by the University of Liverpool Ethics Committee (Project ID 09813).

[†]Berlin School of Economics.

[‡]Technische Universität Berlin, H91, Str. des 17. Juni 135, 10623 Berlin, Germany, michel.tolksdorf@tu-berlin.de. 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. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) and Economic and Social Research Council through NWSSDTP is gratefully acknowledged. Declarations of interest: none.

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.

Ratings have been shown to affect economic interactions. We mainly use them to make better-informed decisions, as ratings can help consumers 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). The prevalence and importance of ratings in economic interactions raise the question about how accurate and, hence, helpful consumer-generated ratings are.

Decades of research in behavioural economics have shown that human decision-making is imperfect and prone to errors and biases. A commonly studied bias is the anchoring effect, which was first proposed by Slovic and Lichtenstein (1972) and further elaborated by Tversky and Kahneman (1974). Anchoring happens through irrelevant informational cues that influence behaviour in a way that is inconsistent with rational decision-making: 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 (Tversky and Kahneman, 1974). If ratings are affected by the anchoring bias, this might hamper their informativeness with detrimental consequences not only for buyers but also for sellers, as well as online market platforms that bring them together. If not accurate, ratings can impair consumers' decision-making and may result in erroneous decisions. Inaccurate ratings might also be harmful to firms; consumers might reward low-quality providing firms and punish more efficient firms – with detrimental effects on overall welfare. Online market platforms are also interested in accurate ratings due to reputation concerns. Hence, it is imperative to understand whether ratings are anchored or not. Potential sources for anchoring could be invoked by the design of the rating system, ratings by other people, and ratings for different products.

Using economic incentives and repeated decisions, this study experimentally scrutinises the prevalence and persistence of the anchoring bias in online ratings by isolating the post-purchase and consumption rating decision. Our experiment has three main features to address the peculiarities of rating systems. First, to closely resemble the provision of ratings in an online environment, we implement an online experiment where participants make rating decisions. 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. Using non-numerical anchors remedies a loss of control arising due to the experiments being conducted online, while the online implementation increases the external validity of our findings. Third, we consider the social component of rating systems by varying whether the anchor is given exogenously or derived endogenously with respect to decisions of group members.

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. Additionally, we implement an unframed control condition, i.e. a slider task. This serves as a rational benchmark, controls for the impact of input errors, and ensures that the framing of our main task works. Lastly, we explicitly control for the impact of cognitive ability and statistical aptitude, as Hoffart et al. (2019) found that the latter affects rating behavior.

The comparison with the slider task 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. 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 cast doubt on the informativeness of ratings and can help to design less error-prone rating platforms that avoid anchoring. We contribute to the literature in two ways. Firstly, we further the knowledge of the existence of anchoring effects in rating settings by providing controlled experimental evidence. Secondly, to our knowledge, we are the first to show the impact of non-numerical anchors on economic decisions.

2 Literature Review

The anchoring bias has been shown to be pervasive and robust in many situations, from consequential economic decisions such as credit card minimum repayments (McHugh and

Ranyard, 2016), auctions (Chui et al., 2022), asset value assessments (Ünveren and Baycar, 2019), real estate evaluations (Northcraft and Neale, 1987) to strategic interactions (Ivanova-Stenzel and Seres, 2021, 2022) and bargaining (Kimbrough et al., 2021) (see Furnham and Boo (2011) for a comprehensive literature review).[1]

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 closely related to the willingness-to-pay for goods, our study is aligned with this literature.

All these studies use exogenous anchors in autonomous and individual decision-making environments. However, many economic decisions take place in social settings. In fact, 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 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. Hereby, the average price of the previous round serves as an anchor. The authors find evidence of significant anchoring effects. Using an estimation task with market conditions such as economic incentives and possible learning through feedback, Meub and Proeger (2015) present robust social anchoring effects. The average estimation of the previous round serves as the anchor. They compare the socially derived anchor to a classical anchor and show that the social context increases the anchoring bias, while they only find weak learning effects. de Wilde et al. (2018) obtain comparable results.

Ratings are provided in social environments. Social information about how other people have rated a certain service or product is available and this information, given accurate, can improve decision quality. But if affected by the anchoring bias, ratings can be detrimental for consumers, sellers, and online market platforms. Research on rating provision has uncovered a J-shaped distribution of ratings (Hu et al., 2009). This translates into some low ratings

[1] 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).

and a lot of high ratings with almost nothing in between these two extremes. Consumers usually rate to 'brag and moan' (Lafky, 2014). They reward or punish sellers depending on whether expectations are met or not and rate leniently under uncertainty (Nosko and Tadelis, 2015; Zervas et al., 2017; Bolton et al., 2019). The stark polarisation and upward compression could lead to anchoring effects. Assuming ratings are strongly skewed towards the positive extreme, this might induce lenient rating behaviour. After observing the positive skewness, an individual who would have provided a medium rating might leave more favourable feedback, which does not reflect the true rating. Besides positive skewness, average ratings, which are a prominent feature on many rating platforms, can also serve as an anchor. For example, Moe and Trusov (2011) empirically identify bandwagon behaviour in ratings. Moreover, Coker (2012) provides evidence of asymmetrical affective perseverance when forming attitudes. Consumers overshoot their judgments when positive information is replaced with negative information but not vice versa. This means positive reviews anchor positive attitudes, even if negative reviews follow later and consumers know previous information was erroneous. Evidence on behaviour that is indicative of the anchoring bias is not only provided for product ratings but also for 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. Berger and Daumann (2021) report anchoring effects for judging basketball players.

Most research that investigates the anchoring bias, including studies on social anchors, heavily rely on abstract, numerical anchors. However, diverging from the standard literature, in this paper, we rely on non-numerical, visual anchors. We do this for two reasons: Firstly, in a rating environment, a lot of information is not only displayed numerically but (sometimes exclusively) visually via coloured stars or bars, smileys, emojis (thumbs up/down), or color-coded indicators. For instance, on Amazon or Uber a continuous five-star rating system is 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. Psychologists discovered some evidence on similar non-standard ways of anchoring. LeBoeuf and Shafir (2006) report non-numerical anchoring effects, using various stimuli such as length, weight, or volume. Oppenheimer et al. (2008) provide experimental evidence for non-numerical and cross-modal anchoring effects. 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, which our study includes.

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	Slider	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 an individual decision-making task where participants provide quality ratings in the rating task conditions. Participants are presented with a slider which represents the overall range of quality. On this slider, a quality interval is displayed. The rating task consist of giving an estimate of the quality that is contained in the quality interval by moving the slider. The task is repeated for twelve rounds without feedback between rounds. There are no numerical values displayed on the slider or quality interval. 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. In three 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, we employ a standard slider task with a neutral framing, meaning that the task is not framed as the rating task. Instead, in the slider task participants are instructed to move the sliders' handle to an invisible target value which is always positioned in the middle of the interval. The closer the participants are to the true quality in the rating task or the target value in the slider task, the higher is the payoff. The participants know this. Thus, the slider task ensures that participants can accurately utilise the sliders' handle. Furthermore, we can identify participants who are not affected by the framing in the rating task conditions but behave similarly to those in the No Framing condition. This ensures that our framing of the rating task works and serves as the rational benchmark.

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. Hereby, 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 rational strategy is to choose the expected quality value which is the middle of the quality interval to minimise deviation from the true quality. In other words, the rational rating is $r^* = \frac{q^H - q^L}{2}$ where q^H is the upper bound of the quality range and q^L is the lower bound of the quality range (see Online Appendix for the formal derivation of the rational rating.)

The displayed quality intervals were randomly drawn in the first session and kept the same across all other sessions to ensure comparability of experimental conditions. To ensure the irrelevance of anchor values, both the maximum and minimum of the slider were excluded from the draw. Subsequently, these values were used as anchor values in the High Anchor and Low Anchor conditions.

In the High and Low Anchor conditions, the participants can either choose the presented anchor as their rating or choose to adjust the rating. The presented anchor is the maximum of the slider in the High Anchor condition and the minimum in the Low Anchor condition. If participants choose to adjust the rating, they must reinitialise their rating by clicking anywhere on the slider. These two anchor types present irrelevant informational cues as the quality interval never contains the minimum or maximum of the slider. In contrast to the High and Low Anchor conditions, the anchor is derived endogenously in the Social Anchor condition and not provided exogenously. 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 adjust their rating. For instance, if the participants are in round two, they see the quality interval and the average rating across all group members from round one as the anchor.

The socially derived anchor does not contain any additional information 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. Participants always have to initialise the slider by clicking on it and can readjust their rating by dragging the sliders' 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 slider task, we collected several control measures. 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 as Hoffart et al. (2019) illustrated that statistical aptitude affects rating behaviour. 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. In the anchoring conditions, we additionally elicited the perceived relevance of the presented anchor by using a 7-point Likert scale. The participants were asked to rate perceived helpfulness, perceived informativeness and cue use from 'I do not agree at all' to 'I fully agree'. Before the experiment started we provided the participants with detailed instructions and control questions to ensure comprehension of the task.[2]

3.1 Participants

Experiments were programmed using oTree (Chen et al., 2016) and conducted virtually in April, May, and October 2021 using the ORSEE database and participant pool of the joint lab of WZB Berlin And TU Berlin (Greiner, 2015). 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. In total 246 participants took part in the experiment over five sessions per experimental condition.

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. Moreoever, 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, it was not unusual for participants to be reminded that they are in "round [one, ..., twelve] out of twelve" solely in written words. While our number of rounds per participant is slightly lower than the 15 rounds used in Meub and Proeger (2015) and Meub and Proeger (2016), Burlig et al. (2020) argue that benefits to power when increasing panel length begin to erode quickly once surpassing short panel lengths of 5 or less.

[2]See Online Appendix for the post-experimental questionnaire, as well as the full translated and original instructions.

3.2 Hypotheses

In our experimental setup, the anchors are designed to be useless and contain no relevant informational value. This means, if participants are not anchored, their rational choice is to select the middle of the quality interval as the rating to minimise deviations from the true quality and maximise payoffs. But as the anchoring effect is prevalent in many economic environments, we expect there to be anchoring effects. The conceptual framework in Figure 1 depicts how we isolate these expected anchoring effects from other confounding factors with our experimental design.

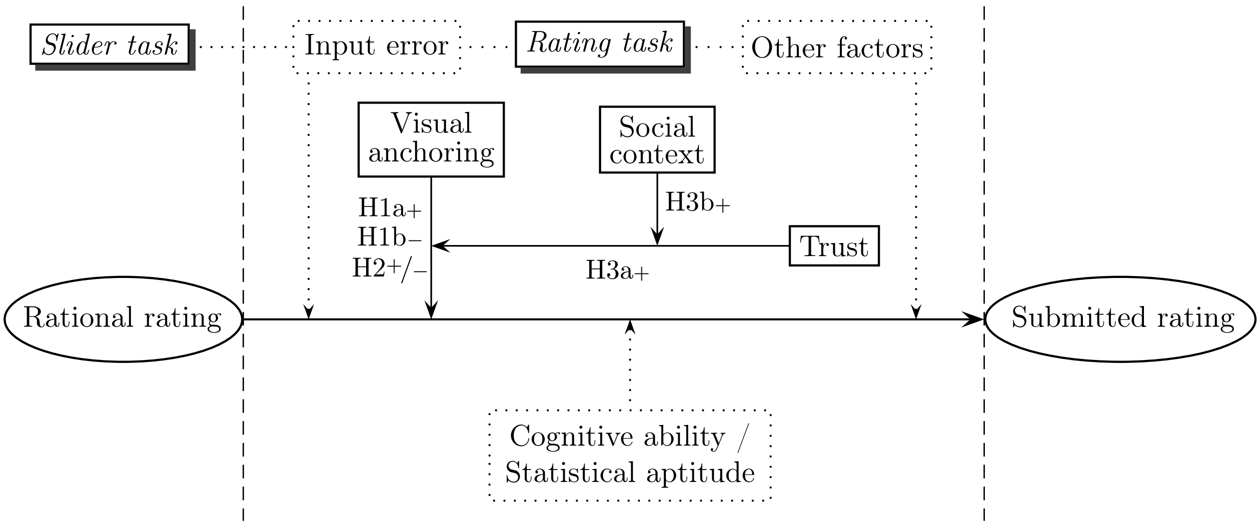


Figure 1: Conceptual framework.

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 the same rating decision for twelve periods. 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 it has been reported, such 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 those 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.*

3.3 Experimental Results

In the following, we explore our hypotheses at hand of the experimental results. We mainly refer to the effects of experimental conditions and control variables on the normalised rating $(r_{it} - r_t^*)/(q_H - q_L)$ or the absolute rating r_{it}. We drop those observations where participants directly chose the presented anchor. Such decisions may come down to one of two reasons. First, participants may have defaulted to the presented anchor. Second, participants may have made this decision by mistake, as it was irreversible. We drop 122 such observations from the Rating Task conditions (6 in High Anchor, 14 in Low Anchor, 6 in Social Anchor outside the quality range, and 42 in Social Anchor inside the quality range). Additionally, we drop one clear outlier in Slider Task where a participant gave a rating that was outside the quality range and 297px from the target while all other ratings are within 21px of the target. This leaves us with a total of 2,829 rating observations by 246 participants. All presented results are robust to keeping the dropped observations.

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 2: Summary statistics of normalised ratings by condition.

In Table 2, 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 from the rational expectation the most, whereas the High and Low Anchor conditions divert somewhat compared to the No Anchor condition. 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 to the No Anchor control.

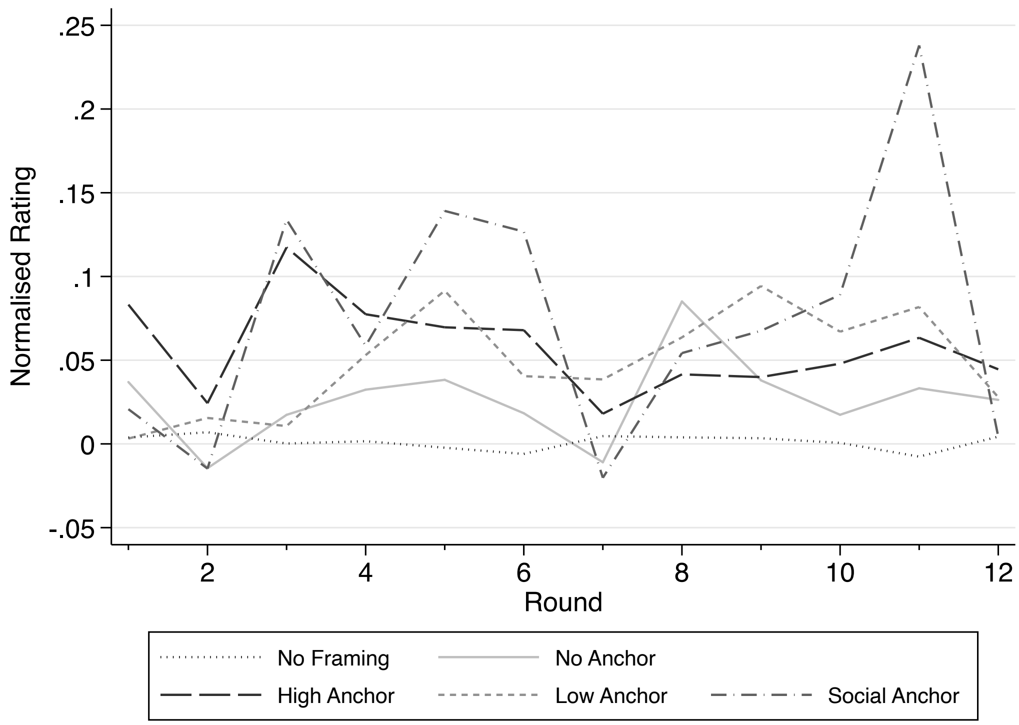


Figure 2: Average normalised ratings over rounds by experimental condition.

In Table 3, we show results 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 Slider Task in the No Framing control, indicating that our framing is effective. When employing a dummy variable indicating the second half 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. Furthermore, it seems that overrating is initially much more pronounced in the High Anchor condition but decreases by the second half of the experiment, as indicated by the marginally insignificant ($p = 0.077$) interaction of High Anchor

	Dependent variable: Normalised rating $(r_{it} - r_t^*)/(q_H - q_L)$			
	(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)
Base category	No Framing	No Framing	No Anchor	No Anchor
Observations	2829	2829	2242	2242
Number of subjects	246	246	197	197

Estimation by OLS regression with standard errors clustered on subject-level in parentheses. Controls are the CRT and statistical aptitude scores. *, **, and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

Table 3: Treatment effects on normalised rating.

and Second half.

Specifications (3) and (4) mirror the approach from before relative to the No Anchor condition, i.e. only for conditions under the Rating Task. 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 and Social Anchor conditions 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 effect of the Low Anchor condition is even positive. The same is true for the interaction of Low Anchor and Second half, indicating that participants might have even over-corrected in the opposite direction of the presented anchor.

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 rational expectation. 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 model explains rating r_{it} of participant i at time t in terms of the rational expectation r_t^* and the anchoring deviation $r_t^* - \bar{r}_{t-1}$ at time t:

$$r_{it} = \alpha r_t^* + \beta(r_t^* - \bar{r}_{t-1}) + \mathbf{x_i}\boldsymbol{\phi} + \epsilon_{it}. \tag{1}$$

For a fully rational participant, we would expect $\alpha = 1$ and $\beta = 0$. That is, the rating is fully described by the rational expectation. An anchoring deviation, i.e. a biased rating towards the anchor is present when $\beta < 0$. The parameter vector $\boldsymbol{\phi}$ captures the effects of controls in form of CRT and statistical aptitude scores contained in $\mathbf{x_i}$. Lastly, ϵ_{it} denotes the 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 types 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.

	Dependent variable: Rating r_{it}		
	(1)	(2)	(3)
Rational rating r_t^*	1.030***	1.007***	1.033***
	(0.0154)	(0.0203)	(0.0237)
Deviation ($r_t^* - \bar{r}_{t-1}$)	-0.0788***	-0.0178	-0.0965**
	(0.0173)	(0.0432)	(0.0331)
F-statistic ($\alpha = 1$)	3.89	0.11	1.99
Prob. > F	0.0544	0.7391	0.1650
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$
Observations	462	193	269
Number of subjects	50	50	50

Estimation by OLS regression with standard errors clustered on subject-level in parentheses. Controls are the CRT and statistical aptitude scores. ** and *** denote significance at the 1% and 0.1% level, respectively.

Table 4: OLS regression of equation (1).

We show the regression results of the aforementioned model in Table 4. 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 rational ratings.

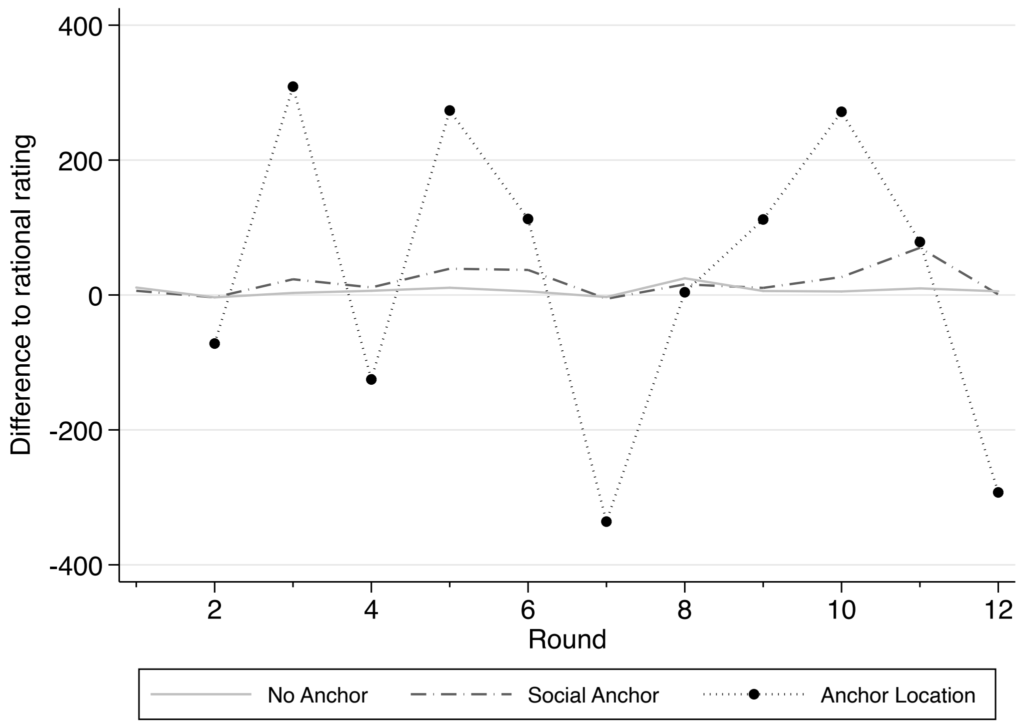


Figure 3: Relation of social anchor location and mean ratings.

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 the social anchors when they are high. The upwards trends do not faint when subsequent anchors fall. Only low social anchors pull ratings down close to the level of the rational predictions. 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.

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 "Mid". In Table 5, we show that by this we classify rounds 3, 5, 6, 9, 10 and 11 as High, rounds 2, 4, 7 and 12 as

Round	1	2	3	4	5	6
SA classification	No	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	Mid	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 5: Classification of Social Anchor (SA) types and calculation of difference between mean normalised ratings in Social Anchor condition and No Anchoring (NA) condition.

Low and round 8 as Mid. Only round 8 is depicted as Mid, 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 where we use an indicator for the "marked rounds" (with a high social anchor present) as an explanatory variable and find that participants significantly overrate in the marked rounds compared to the No Anchor condition. The results are shown in Table 6. 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 and High Anchor conditions in the 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

	Dependent variable: Normalised rating $(r_{it} - r_t^*)/(q_H - q_L)$			
	(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)
Base category	No Framing	No Framing	No Anchor	No Anchor
Observations	1099	2829	1097	2242
Number of subjects	99	246	99	197

Estimation by OLS regression with standard errors clustered on subject-level in parentheses. Marked rounds are rounds 3, 5, 6, 9, 10 and 11. Controls are the CRT and statistical aptitude scores. *, ** and *** denote significance at the 5%, 1%, and 0.1% level, respectively.

Table 6: Impact of high socially derived anchors.

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 and Low Anchoring 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.

Trust is significantly higher in Social Anchor (M = .374, SD = .043) compared to the High Anchor condition (M = .192, SD = .048), $t(97) = 2.853, p = .003$ and compared to the Low Anchor condition (M = .253, SD = .051), $t(97) = 1.842, p = .034$, based on one-sided t-tests. We employ one-sided t-tests as we hypothesised more trust in the Social Anchor condition. There is no significant difference between the High and Low Anchor conditions

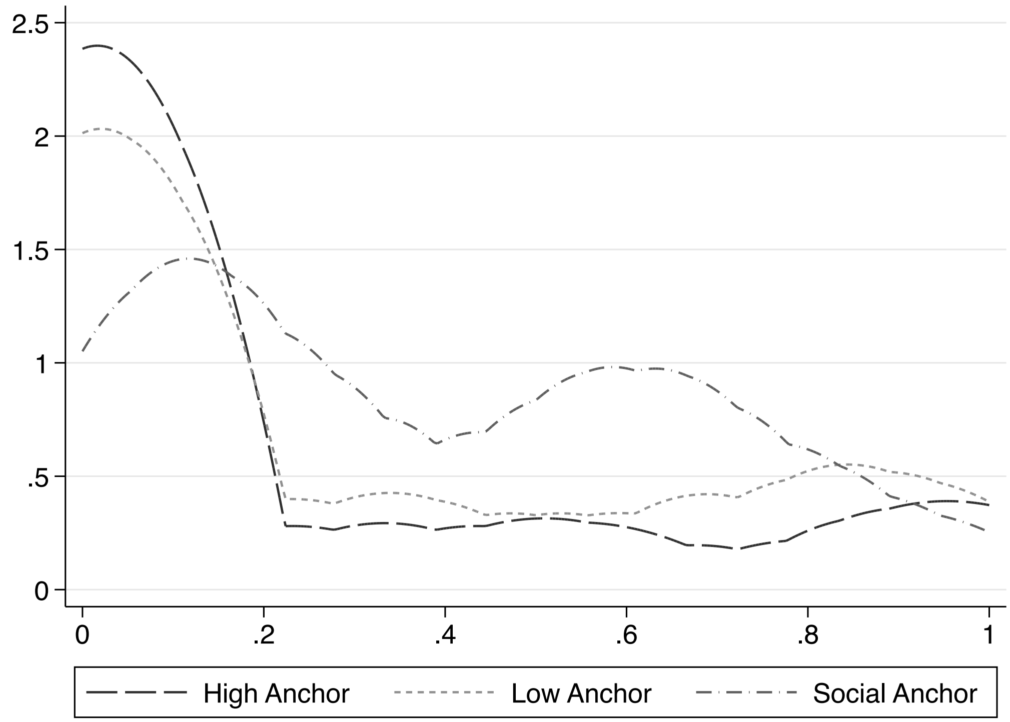


Figure 4: Kernel density estimation of trust score for anchoring conditions.

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 model in (1) by an interaction of the normalised individual trust score s_i with the anchoring deviation $(r_t^* - \bar{r}_{t-1})$:

$$r_{it} = \alpha r_t^* + \beta(r_t^* - \bar{r}_{t-1}) + \gamma s_i(r_t^* - \bar{r}_{t-1}) + \mathbf{x_i}\boldsymbol{\phi} + \epsilon_{it}. \tag{2}$$

Like before, we would expect $\alpha = 1$, $\beta = 0$, and $\gamma = 0$ when participants act rationally. 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 expect $\beta < 0$, $\gamma < 0$, and $\beta + \gamma < 0$.

In Table 7, we show the results of our estimation. Introducing the trust score s_i reveals that we can neither sustain a significant $\beta < 0$, nor do we find a significant $\gamma < 0$, albeit both parameters having a negative sign. However, we find that the joint test $\beta + \gamma < 0$ is significant. This indicates that the priorly depicted anchoring deviation is largely, but not solely, driven by those who place high trust in the signal.

Result 3: *The anchor is perceived as more relevant in the Social Anchor condition com-*

	Dependent variable: Rating r_{it}		
	(1)	(2)	(3)
Rational rating r_t^*	1.030***	1.008***	1.031***
	(0.0155)	(0.0201)	(0.0244)
Deviation $(r_t^* - \bar{r}_{t-1})$	-0.0380	0.00365	-0.0378
	(0.0264)	(0.0526)	(0.0434)
Trust $\times$ deviation $s_i(r_t^* - \bar{r}_{t-1})$	-0.111	-0.0588	-0.151
	(0.0604)	(0.0644)	(0.0975)
F-statistic $(\alpha = 1)$	3.78	0.15	1.66
Prob. > F	0.0578	0.7018	0.2034
F-statistic $(\beta + \gamma = 0)$	11.91**	1.01	6.51*
Prob. > F	0.0012	0.3190	0.0139
Considered observations	All	$\bar{r}_{t-1} < r_t^*$	$\bar{r}_{t-1} > r_t^*$
Observations	462	193	269
Number of subjects	50	50	50

Estimation by OLS regression with standard errors clustered on subject-level in parentheses. Controls are the CRT and statistical aptitude scores. ** and *** denote significance at the 1% and 0.1% level, respectively.

Table 7: OLS regression of equation (2).

pared to the High Anchor and Low Anchor conditions. Trust partially explains the anchoring deviation in the Social Anchor condition.

4 Discussion and Conclusion

In this paper, we have presented an online experiment that studies the prevalence of anchoring effects in a rating environment. We isolated the post-purchase and post-consumption provision of ratings. As ratings are ubiquitous and have become an essential metric guiding our everyday decisions, from what products to purchase or what doctors to consult, it is important to understand whether ratings are anchored by visual cues, suggested ratings, or irrelevant ratings. Anchored ratings can hamper the informativeness of ratings and can 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. Also, firms and online market platforms can suffer reputation damages.

Our online experiment focuses on non-numerical and social anchoring under market conditions such as economic incentives and repeated decision-making. The study features three different anchors, which are low, high, and socially derived anchors. We uncovered significant anchoring effects. For all anchoring conditions, we observe rating inflation with the high and social anchors having the highest impact. The effect of the high anchor is significantly persistent throughout. But the anchoring effect is asymmetric as for the low anchor there is no significant anchoring effect. The socially derived anchor takes on varying roles between a high, a low and a mid-anchor and is indeed directly predictive of the observed ratings. Our overall findings do not support the notion that market conditions act as a filter for heuristics and biases as we observe anchoring effects despite economic incentives. We observe anchoring effects in a rather simple environment, that are not driven by numbers, only through visual changes. If anchoring effects are prevalent in non-numerical settings such as ours, we could expect similar anchoring effects in many real-world rating systems, especially those implemented in online markets and platforms.

When ratings provided in each anchoring condition are compared to a slider task which is almost identical to the rational benchmark, we observe significant overrating in all anchoring conditions. This is consistent with other studies that scrutinise rating behaviour under uncertainty. Clients tend to provide lenient ratings, which in turn can exacerbate the upward compression of ratings. This is also a more general pattern in rating behaviour that has been uncovered previously, with usually a large fraction of positive feedback with very few negative ratings. This pattern is also mirrored by asymmetric anchoring effects. Whereas it is possible to inflate the ratings, low anchors have no significant effects on ratings. Our study indicates that irrelevant informational cues that work as anchors can contribute to the positive skewness of ratings.

Our results show how easy it is to bias and inflate ratings, as they are prone to high anchors. Many online market platforms rely on truthful ratings as part of their marketing strategy to create and promote an accurately reflected reputation. But our results highlight that market platforms should not solely rely on ratings but also include other factors, such as the number of explicit complaints and returns. At first glance, rating inflation seems to be favourable for firms. However, upward compression of ratings makes it harder for firms to set themselves apart from their competitors through positive ratings. It also makes it more difficult to distinguish between good-quality and bad-quality offering firms. Furthermore, anchored ratings make it more difficult for firms to use clients' ratings as feedback to assess the quality of their product.

Our findings are related to Sugden et al. (2013), as we employ plausible anchors and, similar to them, uncover asymmetric anchoring effects. We find that only high anchors, both social and non-social, are effective, whereas low anchors are not. Like de Wilde et al. (2018), we find that the social anchor is perceived as more relevant even though it does not contain any informational value in our experiment. There are two potential explanations for this.

Firstly, as each participant contributes into the social anchor, this might create a notion of overconfidence where each participant overweighs their own rating which in turn increases perceived importance and trust in the social anchor. Secondly, the social anchor might create an illusion of the wisdom of crowds inducing more trust in the endogenously derived anchor. Which of these two explanations hold, is subject to future research.

Overall, our study contributes to two strands of literature. Firstly, we contribute to the literature on ratings and platform design in that we extent the knowledge about anchoring effects to the domain of rating settings in form of controlled experimental evidence. Our study can help to design less error-prone rating platforms where anchoring is avoided. Secondly, we add to the literature on non-standard anchors by focusing on social and non-numerical anchors. This contributes to the anchoring literature which, hitherto, strongly focuses on numerical anchoring. While we focused on the opposite case of purely non-numerical anchoring, notably, many rating environments in the real-world are a hybrid of both, e.g. a 5-star rating system. A promising future avenue for research could be to focus on non-numerical but countable rating systems.

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.

Berger, T. and Daumann, F. (2021). Anchoring bias in the evaluation of basketball players: A closer look at nba draft decision-making. Managerial and Decision Economics, 42(5):1248–1262.

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.

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.

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.

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.

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.

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.

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).

Zervas, G., Proserpio, D., and Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. Journal of Marketing Research, 54(5):687–705.

Non-numerical and social anchoring in consumer-generated ratings

Online Appendix

Yigit Oezcelik[*], Michel Tolksdorf[†‡]

December, 2023

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

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

[*] Corresponding author. University of Liverpool, University of Liverpool Management School, Mulberry Court, Mulberry Street, Liverpool L69 7ZH, United Kingdom, Yigit.Oezcelik@liverpool.ac.uk.

[†] Berlin School of Economics.

[‡] Technische Universität Berlin, H91, Str. des 17. Juni 135, 10623 Berlin, Germany, michel.tolksdorf@tu-berlin.de.

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

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

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]

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]

Reset

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

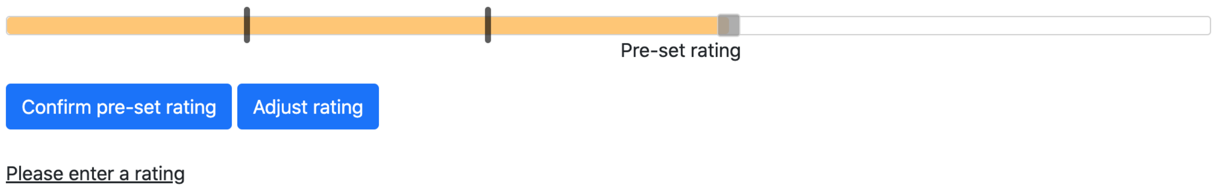
Reset

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

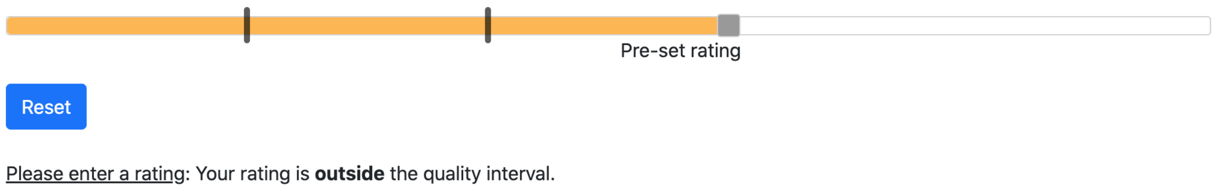
[End examples]

[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.

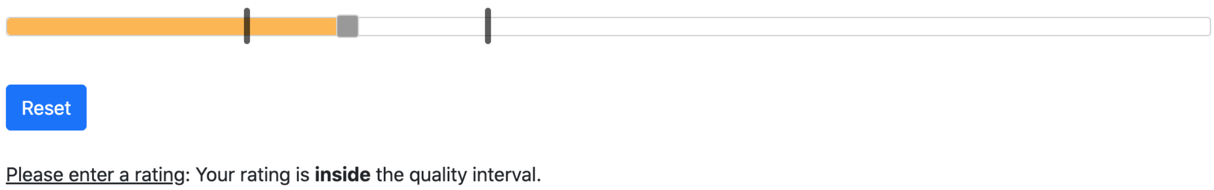
Your task consists of rating the quality you obtain. You can provide a rating with the help of the handle on the bar. First, you can choose whether you want to choose a pre-set rating or adjust the rating. If you choose the pre-set rating in a round, you cannot revoke it. If you want to adjust the rating, 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

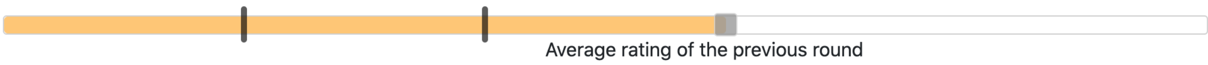
[Example of interaction]

Reset

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

[End example]

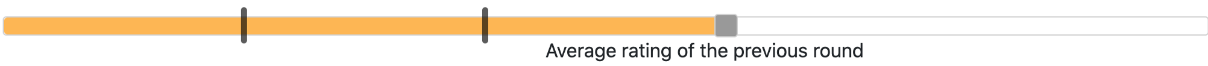
Beginning with the second round, the average rating of your group (including your rating) from the previous round will be displayed next to the quality interval. First, you must decide to either choose the average rating of the previous round or to adjust the rating. If you choose the average rating 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.



Choose the average rating of the previous round | Adjust rating

Please enter a rating

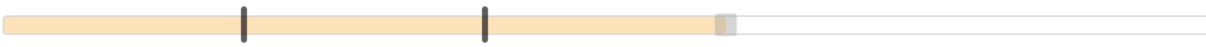
[Example of interaction after 'Choose average rating of the previous round']



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.

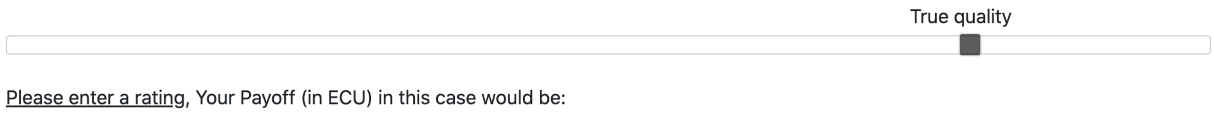
[End examples]

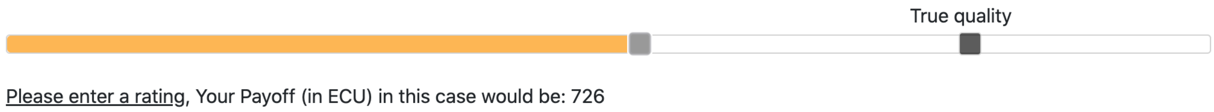
[End conditions]

[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.

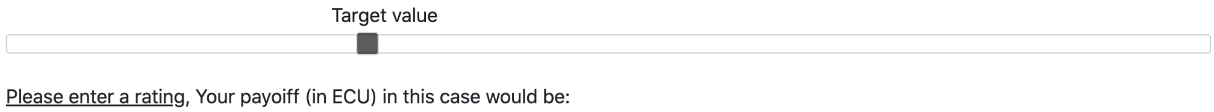


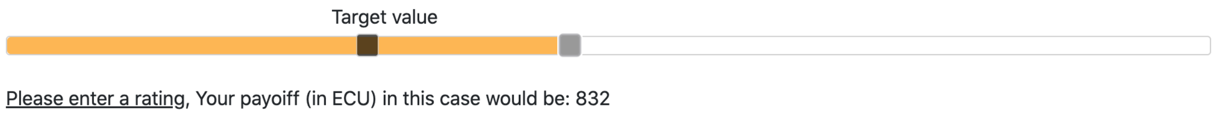


[No Framing condition]

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.





[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

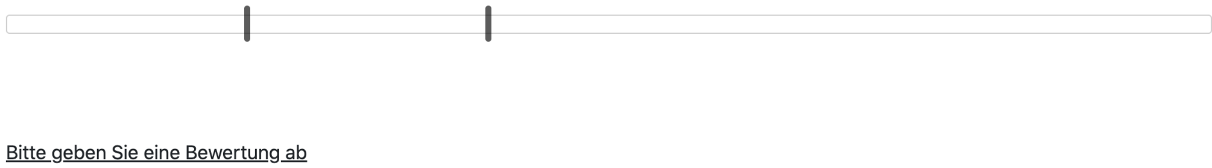
[Text in brackets was not observed by subjects. Presented sliders were interactive in the digital instructions. For examples of the interactions see translated instructions.]

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.

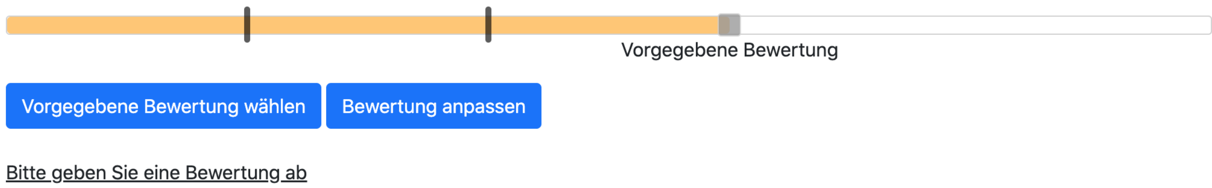
[No Anchor condition]
Ihre Aufgabe:
In diesem Experiment wird Ihnen jede Runde ein Qualitätsintervall angezeigt, welches auf einem Balken liegt. Die wahre Qualität liegt in diesem Intervall, wobei jeder Wert im Intervall gleichwahrscheinlich ist. Die Qualität steigt auf dem Balken von links nach rechts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



[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.



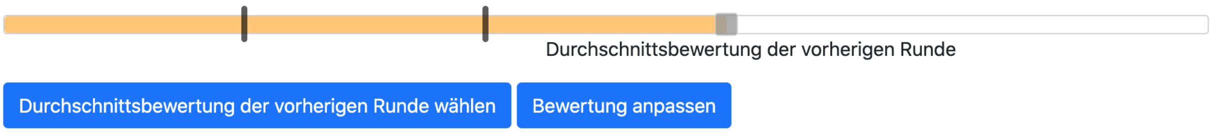
[Social Anchor condition]
Ihre Aufgabe:
In diesem Experiment werden sie einer Gruppe zugeordnet, welche aus Ihnen und vier weiteren Teilnehmern besteht. Diese Gruppe bleibt während der zwölf Runden des Experiments gleich und ändert sich nicht. Ihnen und allen anderen Gruppenmitgliedern wird jede Runde 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.

Bitte geben Sie eine Bewertung ab

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.



Bitte geben Sie eine Bewertung ab

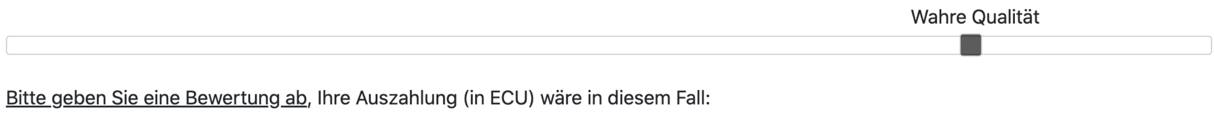
[No Framing condition]
Ihre Aufgabe: In diesem Experiment wird Ihnen jede Runde ein Intervall angezeigt, welches auf einem Balken liegt. Genau in der Mitte innerhalb des Intervalls ist der Zielwert.

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.

Bitte machen Sie eine Eingabe

[End experimental conditions]

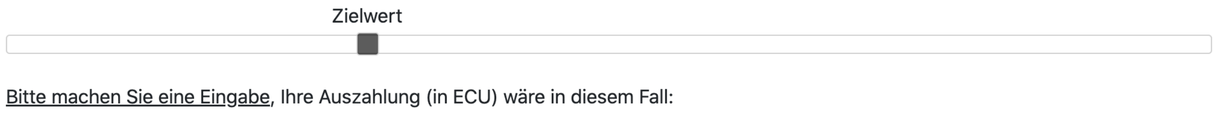
[No Anchor, High Anchor, Low Anchor and Social Anchor conditions]
Ihre Auszahlung:
Insgesamt gibt es zwölf Runden. In jeder Runde ist das Qualitätsintervall und damit die wahre Qualität unabhängig von der vorherigen Runde. Nach Abschluss der zwölf Runden wird zufällig eine Runde gewählt, welche Ihre 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.



[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.



[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]
Review Questions:
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]
Review Questions:
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]