



# On the effect of anchoring on valuations when the anchor is transparently uninformative

Konstantinos Ioannidis<sup>1</sup> · Theo Offerman<sup>1</sup> · Randolph Sloof<sup>2</sup>

Received: 3 December 2019 / Revised: 21 July 2020 / Accepted: 18 August 2020 /  
Published online: 22 September 2020  
© The Author(s) 2020

## Abstract

We test whether anchoring affects people’s elicited valuations for a bottle of wine in individual decision-making and in markets. We anchor subjects by asking them if they are willing to sell a bottle of wine for a transparently uninformative random price. We elicit subjects’ Willingness-To-Accept for the bottle before and after the market. Subjects participate in a double auction market either in a small or a large trading group. The variance in subjects’ Willingness-To-Accept shrinks within trading groups. Our evidence supports the idea that markets have the potential to diminish anchoring effects. However, the market is not needed: our anchoring manipulation failed in a large sample. In a concise meta-analysis, we identify the circumstances under which anchoring effects of preferences can be expected.

**Keywords** Anchoring · Replication · Market · Experiment · Meta-analysis

**JEL Classification** D01 · D91 · C91

## 1 Introduction

A wealth of evidence has accumulated questioning some of the foundations of expected utility theory, and behavioral theorists have shown how these challenges can be accommodated (Wakker 2010). At the core of standard and behavioral economic modelling remains the assumption that people are endowed with

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40881-020-00094-1>) contains supplementary material, which is available to authorized users.

---

✉ Theo Offerman  
[t.j.s.offerman@uva.nl](mailto:t.j.s.offerman@uva.nl)

Konstantinos Ioannidis  
[ioannidis.a.konstantinos@gmail.com](mailto:ioannidis.a.konstantinos@gmail.com); [k.ioannidis@uva.nl](mailto:k.ioannidis@uva.nl)

<sup>1</sup> CREED, University of Amsterdam and Tinbergen Institute, Amsterdam, The Netherlands

<sup>2</sup> University of Amsterdam and Tinbergen Institute, Amsterdam, The Netherlands

well articulated and stable preferences. This fundamental assumption, however, has been challenged amongst others by Ariely et al. (2003), who have shown that preferences are initially malleable by normatively irrelevant anchors. People subsequently choose consistently with these initial preferences, and thereby end up with preferences that are characterized by what Ariely et al. (2003) call “coherent arbitrariness”. For a series of products that range from familiar (like an average bottle of wine) to unfamiliar (like listening to an unpleasant sound), they find substantial anchoring effects.

Economists often assign less weight to behavioral anomalies when they are obtained in non-repeated individual decision-making tasks. The line of reasoning is that anomalies may be eroded when people have relevant experience, for instance, as a result of trading in markets. To counter such skepticism, Ariely et al. (2003) included a treatment where subjects, after being exposed to an anchor, submitted a bid to avoid listening to an annoying sound. In the uniform-price sealed-bid auction, the three lowest bidders had to listen to the sound and each of them received a payment equal to the fourth lowest bid. Like in the individual decision-making treatment, sizable (and lasting) anchoring effects were observed in this treatment.

This paper aims to make two contributions. A first contribution is that we investigate the effects of uninformative anchoring on valuations for a familiar good in a large sample. This is important because previous papers have provided mixed evidence, from sizable anchoring effects (Ariely et al. 2003) to no anchoring effects (Fudenberg et al. 2012). Our paper stands out because of the combination of two features. First, we have a large sample of 316 subjects who are all exposed to the same anchoring protocol, while previous studies have often been based on rather small samples. Second, we use a transparently random anchor that subjects know to be uninformative because they generate it themselves with a ten-sided die.

A second contribution of our paper is that we investigate how elicited preferences are affected in a richer market setting than the one of Ariely et al. (2003), where subjects could not learn from others’ bids during the auction. We employ a standard double auction where traders are continuously updated about other traders’ bids and asks. We believe that a double auction provides a much better chance for market forces to erode initial traces of anchoring.

Our experiment consists of three phases. In the first phase, we apply a typical anchoring protocol: we ask whether subjects are willing to sell a bottle of wine for an individually drawn, random price. Then we elicit their valuation (Willingness-To-Accept) for the bottle of wine with the Becker–DeGroot–Marschak (BDM) procedure (Becker et al. 1964). In the second phase, we randomly assign subjects to either a small double auction market ( $n = 2$ ) or a large double auction market ( $n = 8$ ). Subjects participate in two trading periods, once as a buyer and once as a seller. In the third phase, we elicit each subject’s valuation once more.

The first phase of the experiment allows us to test whether a random anchor influences elicited valuations. We hypothesize that subjects’ valuations correlate positively with their anchors. We further conjecture that market experience will affect subjects’ elicited preferences. Subjects who are not completely sure about their preference may move into the direction of the preferences exhibited by other traders. This way, anchoring effects may diminish or even disappear. Thus, we hypothesize

that the valuations elicited in the third phase will exhibit smaller (if any) anchoring effects. We also hypothesize that the large market will have a stronger effect on subjects' preferences than the small market, and that anchoring effects are eroded more efficiently in the former.

Contrary to our first hypothesis, we observe no effect of the random anchor on subjects' valuations. We believe our null result contributes to the literature on the robustness of anchoring effects. In Sect. 4, we position our paper in the literature and elaborate on what we can learn from our null result. There, we discuss the results of a concise meta-analysis of experimental papers that cite Ariely et al. (2003) and investigate the effects of anchoring on preferences.

We do find support for the idea that market participation affects how people value the bottle of wine. The variance in subjects' elicited valuations after the market shrinks within trading groups. As expected, the effect of other traders' behavior on a subject's preference is stronger in the large market. These results underline the potential power that markets may play in eroding individual biases and noise. However, in this study, the double auction is not needed to avoid anchoring effects on valuations.

The remainder of the paper is organized as follows. Section 2 describes our experimental design and the hypotheses to be tested. Section 3 presents the results of the experiment. Section 4 provides a discussion of how our results fit in the literature.

## 2 Experimental design and implementation

We pre-registered our study on the American Economic Association's registry for randomized controlled trials (Ioannidis et al. 2018).<sup>1</sup> The experiment was run at the CREED communication Lab of the University of Amsterdam. The communication lab has 16 soundproof, closed cubicles. The experiment was programmed in oTree (Chen et al. 2016). Subjects read the computerized instructions at their own pace (see supplementary material for instructions). No communication was allowed during the experiment. Subjects were informed that they could earn money as well as a bottle of wine. It was explained that the experiment consisted of three phases during which they would make five decisions. Subjects knew that one of those five decisions would randomly be selected for payment at the end of the experiment. In phase I, subjects made two decisions, in phase II they made two decisions and in phase III they made one decision. They only received the instructions for the next phase after a previous phase was finished.

There were two treatments which were varied between subjects. The Small market consisted of two subjects and the Large market of eight subjects. In each session,

---

<sup>1</sup> If we had found that elicited valuations are affected by anchoring and that markets diminish the role of anchoring, a confounding explanation would be that the effect of anchoring generally fades out over time. Our pre-registration mentions a control treatment to isolate the part of the reduction of the anchoring effect due to market forces and the part due to time fading. Given that we do not find an effect of anchoring, we did not run this treatment.

we simultaneously ran the two treatments. Subjects were randomly assigned to either one of them.

Phase I was identical for both treatments. At the start of phase I, the experimenter entered each subject's cubicle with a ten-sided die (numbered from zero to nine). Subjects determined their own random anchor by rolling the die twice. The first outcome was the integer part and the second was the decimal part of the anchor price. For example, if a subject rolled six and four, the price was 6.4€. Hence, subjects knew that the anchor price was an uninformative draw in the range from 0€ up to 9.9€. This procedure took place in the presence of the experimenter to guarantee that the subjects entered the correct numbers.<sup>2</sup> We used this procedure of subjects generating the anchor themselves to make it fully transparent to our subjects that the anchor price was truly random.

The first decision of phase I was the anchoring question. The subjects were endowed with a bottle of wine, a picture of which was shown to them on their screen. Consequently, they were asked whether they were willing to sell the bottle to the experimenter for a price that corresponded to the anchor price that they had just drawn. For the second decision of phase I, each subject was asked to submit the minimum price for which they were willing to sell the same bottle of wine. This Willingness-To-Accept decision (WTA) was incentivized via the BDM procedure. The application of the BDM procedure aimed at minimizing the chance that subjects form any kind of inference from the elicitation process itself. The instructions included a description of the BDM mechanism and emphasized that it is optimal to provide the true valuation of the bottle. The explanation did not include a numerical example as we did not want any number to operate as an additional anchor. For the same reason, the upper bound of the distribution from which the BDM price was drawn was not revealed. The subjects knew that a number would be randomly drawn between 0 and two times the (unknown) price of the bottle of wine in the store. To avoid outliers, we bounded the WTA from above. Subjects were given an error message if they entered a WTA above two times the price of the wine and were asked to resubmit their decision.<sup>3</sup> The message did not inform them of the actual upper bound, but simply stated that their price was higher than what the experimenters believe is a reasonable price for the wine.<sup>4</sup>

In phase II, the market treatment was implemented. In the Large market, eight subjects participated in a double auction with four buyers and four sellers. In the Small market, two subjects participated in a market with one buyer and one seller. In a typical session of 16 subjects, half were randomly assigned to the Large market (one trading group) and half to the Small market (four trading groups). The market lasted for two periods. The trading group remained the same across the two periods,

<sup>2</sup> Four out of 316 subjects did not wait until the experimenter arrived and entered numbers of their own.

<sup>3</sup> This message was shown to only 5 out of 316 subjects.

<sup>4</sup> Bohm et al. (1997) showed that selling prices elicited via a BDM mechanism are sensitive to the upper bound of the BDM distribution. They use three treatments varying the bound, namely standard (market price), high (unrealistic price) and unspecified (upper bound as "not to exceed what we believe any real buyer would be willing to pay"). They observe no difference between the standard and unspecified, whereas bidding is higher in the high treatment.

but buyer and seller roles were swapped. This way all subjects were exposed to both sides of the trade before they continued to phase III.

Except for the number of traders, the market treatments were identical. Each seller was endowed with a bottle of wine and each buyer was endowed with an amount equal to the price of the bottle of wine that we paid in the store. Traders were unaware of the size of this amount. At the end of the experiment, the amount was revealed only if the market decision was chosen for payment and only to buyers.

Buyers could submit bids to buy the bottle of wine. They could increase their bid multiple times, but not decrease (or withdraw) their current highest bid. Sellers could submit asks to sell the bottle of wine. They could decrease their ask multiple times, but not increase (or withdraw) their current lowest ask. All bids and asks were automatically recorded in the Order Book, which was visible to everyone and updated in real time. A trade occurred automatically whenever any of the following two rules was satisfied. (i) When a buyer submitted a bid that was higher than or equal to the lowest ask of the sellers in the Order Book, this buyer bought from the seller with the lowest ask and the corresponding ask was the transaction price. (ii) When a seller submitted an ask that was lower than or equal to the highest bid of the buyers in the Order Book, this seller sold to the buyer with the highest bid and the corresponding bid was the transaction price. All realized trades and their corresponding prices were automatically recorded in the publicly visible Trade Book. Subjects who had already traded still saw live updated Order and Trade Books. Before the trading period opened, the subjects had to correctly answer six multiple choice questions to make sure they understood the rules of the market.

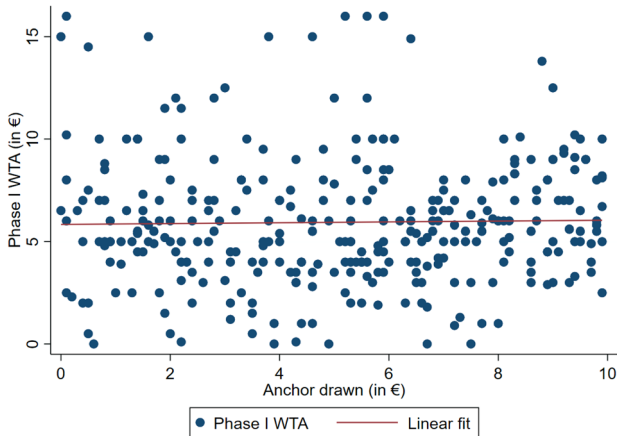
In phase III, we again elicited subjects' WTA for the bottle of wine with the BDM mechanism. After that, subjects were asked to complete a standard demographics survey asking for their age, gender and field of study. The experiment ended at this point and the final screen shown to the subjects informed them about which of the five decisions was chosen for payment as well as their payoff. If a WTA decision was implemented, they were informed of the random BDM draw and whether this random price meant that they sold the bottle of wine or kept it.

This design allows us to test the following hypotheses. To that purpose, we use the anchors to assign subjects to a High-anchor group and a Low-anchor group on the basis of either a median split or a quartile split. We use the data of phase I to test for anchoring.

**Hypothesis 1** The phase I WTA in the High-anchor group is larger than the phase I WTA in the Low-anchor group.

We use the data of phases I and III to test whether the market affects subjects' elicited preferences and alleviates the anchoring effect.

**Hypothesis 2** The difference in WTA between the High and the Low-anchor group is smaller in phase III than in phase I.



**Fig. 1** Scatter plot of phase I WTA on anchor with fitted regression line

**Hypothesis 3** The reduction in the difference in WTA between phase I and III is larger in the Large market treatment than the Small market treatment.

In total, 316 subjects participated in the experiment, 160 in the Large market treatment (20 trading groups) and 156 in the Small market treatment (78 trading groups). The experiment lasted approximately 75 minutes. Depending on their decisions during the experiment, subjects received on average 12.17€ including the participation fee of 8€ (excluding the bottle of wine). On top of their payment, 160 of the subjects physically received a bottle of wine. We used four different bottles of wine across sessions to avoid that prospective subjects could potentially learn the price from subjects that had participated already. Two of the bottles were priced at 6.00€ and two at 7.50€. Our subjects are on average 21 years old. Most of them (67%) are economics students, and they are evenly balanced across genders (females 53%, males 47%).

### 3 Results

#### 3.1 Anchoring manipulation

In this subsection, we shed light on the question whether anchoring affects subjects' valuation of the bottle of wine. Figure 1 plots subjects' WTA in phase I as a function of their anchor. The figure suggests that subjects' WTA is fairly independent of their anchor.<sup>5</sup>

<sup>5</sup> All the results presented in this subsection are robust to discarding the subjects that did not wait for the experimenter to record their anchor (see footnote 2) and the subjects that reported an unreasonable high initial WTA (see footnote 3).

**Table 1** Mean WTA by anchor group

	Phase I WTA		
	Median	Quartile	Quintile
High-anchor group	6.07 (0.23)	6.35 (0.29)	6.74 (0.32)
Low-anchor group	5.79 (0.27)	6.11 (0.37)	6.29 (0.44)
Ratio (High/Low)	1.048	1.039	1.072
$z$	1.314	1.137	1.249
$p$ value	0.189	0.257	0.213
Observations	316	163	123

$z$  and  $p$  values refer to Wilcoxon–Mann–Whitney rank-sum tests. Standard errors in parentheses

Table 1 makes the results more precise. Hypothesis 1 states that the WTA in the group with high anchors will be larger than the WTA in the group with low anchors. First, we do a median split of our data. Contrary to the hypothesis, the WTA for the Low-anchor group does not significantly differ from the High-anchor group. The evidence is in the expected direction, but the effect size is very small and far from economically significant. The magnitudes of our anchoring effects as measured by the ratio of the valuations in the top and bottom part of the distribution varies between 1.04 (for the ratio of the quartiles) to 1.07 (for the ratio of the quintiles).<sup>6</sup> In comparison, for the series of products in Ariely et al. (2003) the ratio of top and bottom quintiles ranges from 2.16 to 3.03. The lack of support for an anchoring effect is further illustrated by a regression of the reported WTA on the anchor, while controlling for the price of the wine. (All regression results reported in the paper are obtained from OLS regressions.) The estimation reveals a very small and far from significant slope ( $b = 0.019$ ,  $SE = 0.061$ ,  $CI = [-0.102, 0.140]$ ,  $t = 0.31$ ,  $p = 0.755$ ,  $N = 316$ ).

The previous literature has suggested some robustness checks. For instance, Fudenberg et al. (2012) include an analysis where they test for anchoring effects after leaving out inconsistent responses. We define a response as inconsistent if the WTA is higher than the anchor price that was accepted or lower than the anchor price that was rejected. In our sample, we have 51 (16.14%) inconsistent observations from subjects resulting in a reduced sample size of 265. Using rank-sum tests, we find no anchoring effect for either median split (ratio = 1.177,  $z = 0.239$ ,  $p = 0.811$ ,  $N = 265$ ) or quartile split (ratio = 1.027,  $z = 0.971$ ,  $p = 0.429$ ,  $N = 135$ ) or quintile split (ratio = 1.022,  $z = 0.738$ ,  $p = 0.460$ ,  $N = 105$ ). A regression of valuation on anchor—again controlling for price—reveals an insignificant slope ( $b = -0.015$ ,  $SE = 0.063$ ,  $CI = [-0.140, 0.110]$ ,  $t = -0.24$ ,  $p = 0.813$ ,  $N = 265$ ). Hence, focusing only on consistent answers does not affect our main result of no anchoring effects.

<sup>6</sup> To have enough observations in each group, we preregistered to run the tests on the top versus the bottom half, and on the top versus the bottom quartile. The literature focuses on quintiles instead of quartiles. For comparison, we have included these statistics as well.

Another approach that has been used in the literature is to replace valuations above the BDM range by the maximum of the BDM range. One reason to do so is that all reports higher than the BDM range yield the same outcome. So very high reports need not reflect very high valuations, which may bias the analysis. Ariely et al. (2003) and Maniadis et al. (2014) truncate valuations in this way and find that it does not affect their results. The same approach is not directly applicable for our study as our subjects did not know the exact range of the BDM, and higher valuations than the maximum were not allowed. However, in the same spirit we can investigate whether our results are sensitive to replacing valuations above 10 by 10, the highest possible anchor. Rank-sum tests reveal no anchoring effects for either median split (ratio = 1.077,  $z = 1.359$ ,  $p = 0.174$ ,  $N = 316$ ) or quartile split (ratio = 1.081,  $z = 1.169$ ,  $p = 0.242$ ,  $N = 163$ ) or quintile split (ratio = 1.177,  $z = 1.278$ ,  $p = 0.201$ ,  $N = 123$ ). A regression of valuation on anchor and price confirms the result ( $b = 0.055$ ,  $SE = 0.051$ ,  $CI = [-0.046, 0.156]$ ,  $t = 1.07$ ,  $p = 0.285$ ,  $N = 316$ ). Hence, the truncation of valuations also does not qualify our null result.

As a final robustness check, we test for anchoring across demographic characteristics of our subjects (field of study and gender) as well as across different types of wine. We regress valuation on anchor, controlling for the market price of the wine and find no anchoring effect both for economic students ( $b = 0.049$ ,  $SE = 0.075$ ,  $CI = [-0.099, 0.198]$ ,  $t = 0.66$ ,  $p = 0.511$ ,  $N = 211$ ) as well as non-economic students ( $b = -0.037$ ,  $SE = 0.106$ ,  $CI = [-0.247, 0.173]$ ,  $t = -0.35$ ,  $p = 0.735$ ,  $N = 105$ ). We repeat the same exercise and find no anchoring effect both for male students ( $b = -0.023$ ,  $SE = 0.089$ ,  $CI = [-0.200, 0.154]$ ,  $t = -0.26$ ,  $p = 0.797$ ,  $N = 148$ ) as well as female students ( $b = 0.061$ ,  $SE = 0.085$ ,  $CI = [-0.108, 0.229]$ ,  $t = 0.71$ ,  $p = 0.477$ ,  $N = 168$ ). Similarly, we find no anchoring effect for any of the four types of wine we used.<sup>7</sup>

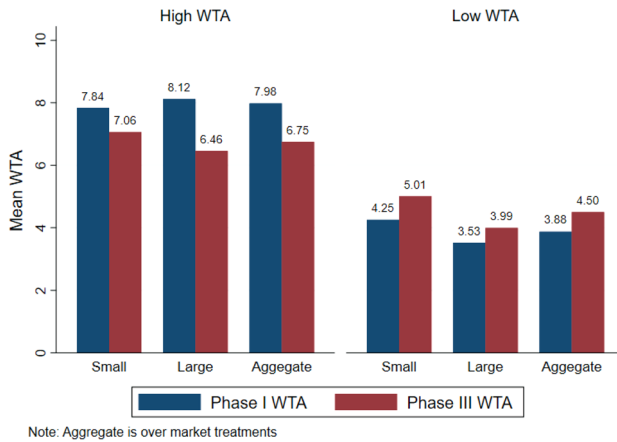
**Result 1** There is no anchoring effect in our data.

Two factors may play a role in our null-result for the effect of anchoring on valuations. The first is the familiarity of the product. Most subjects are likely to be familiar with a bottle of wine, and it may be that anchoring effects occur more easily for unfamiliar products for which people lack a clear initial preference. The other is that in our study, the anchoring procedure is transparently uninformative. In the concluding discussion, we present the results of a small meta-analysis that sheds light on these factors.

In light of Result 1, any analysis on whether market experience reduces anchoring effects is meaningless. Still, it remains interesting to investigate whether the market affects people's preferences. Previous work showed that markets can affect people's preferences for unfamiliar goods for which people might not have a clear

<sup>7</sup> We regress valuation on anchor separately for each type of wine and the coefficient of anchor is never significant. The  $p$  values range from 0.131 to 0.791.





**Fig. 2** Average WTA before and after the market

initial preference to start with, such as tasting an unpleasant liquid (Tufano 2010) and lotteries (Isoni et al. 2016). It is not clear that markets can affect people's preferences for more familiar goods like a bottle of wine. For the remainder of the analysis, we group subjects on the basis of their phase I WTA instead of their anchor and reshape the remaining two hypotheses accordingly.

### 3.2 Market effect on valuations

To investigate how markets affect elicited preferences, we define new groups based on the valuations. For each trading group separately, we use a median split of the phase I WTA to assign each subject to a Low- or High-WTA group.<sup>8</sup>

Hypothesis 2 deals with the question whether the information revealed during the market affects the valuation of subjects. Figure 2 illustrates the results. Focusing on the aggregate results of the Small and the Large markets, it is clear that subjects' valuations move in the direction of the other WTA group.<sup>9</sup>

To test whether the change is a statistical artifact due to regression to the mean, we compare for each subject the absolute difference between their WTA from phase I and the average WTA (from phase I) in their trading group with the same variable from phase III. The average absolute difference is 2.06€ in phase I and 1.64€ in phase III. A Wilcoxon signed-rank test reveals that traders' WTA vary less after the market than before ( $z = -5.321, p < 0.001, N = 316$ ).

<sup>8</sup> If, for example, in the Small market, one subject in a trading group submits a valuation of 1€ and the other a valuation of 2€ we classify the latter in the High-WTA group, even though his WTA is low in comparison to the overall sample of subjects.

<sup>9</sup> Subjects in the Low-WTA group increased their valuation significantly by 0.62€ ( $z = 3.720, p < 0.001, N = 158$ ) and in the high-WTA group decreased it significantly by 1.23€ ( $z = -5.511, p < 0.001, N = 158$ ).

**Table 2** Effect of observed market information on WTA change by treatment

	Small Phase III WTA	Large Phase III WTA	Both Phase III WTA
Phase I WTA	0.64*** (0.08)	0.47*** (0.08)	0.50*** (0.06)
Observed market information	0.33*** (0.08)	0.36* (0.14)	0.34*** (0.07)
Small*Phase I WTA			0.10* (0.05)
Wine price	0.32 (0.24)	0.08 (0.25)	0.25 (0.17)
Constant	1.80 (2.32)	- 2.03 (1.46)	- 0.52 (1.39)
Controls	Yes	Yes	Yes
$R^2$	0.520	0.421	0.470
Observations	156	160	316

Controls are gender, age and field of study. Standard errors in parentheses are adjusted for 98 trading group clusters. All regressions are estimated using OLS.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

**Result 2** Subjects change their WTA in the direction of the average WTA in their own trading group. WTA's elicited in phase I vary more within their trading group than WTA's in phase III do.

We now turn to the question whether subjects change their preferences more in the Large market than in the Small market. Figure 2 also displays the results for each market separately. In agreement with Hypothesis 3, we observe that the average decrease in WTA for the High-WTA group is larger in the Large market than in the Small market. In contrast to Hypothesis 3, subjects in the Low-WTA group increase their WTA to a somewhat larger extent in the Small market than in the Large market.

We test the differential impact of the Large market on preferences in a regression that explains the phase III WTA by the phase I WTA and observed market information, with and without interaction term for the treatment. We define observed market information as the average of the last observed actions of the other market subjects. For market subjects that traded, the last observed action is the price they agreed on. For market subjects that did not trade, it is the last bid/ask they submitted. The results in Table 2 provide supportive evidence for the idea that subjects attach more weight to their own WTA in the Small market as compared to the Large market.<sup>10</sup> In the first two columns, we present regressions for each market separately and in the last column, we present a regression with both markets and an interaction term.

<sup>10</sup> In the regression Table 2, we included the market price of the wine. Using wine fixed effects instead produces qualitatively similar results.

**Result 3** Subjects change their valuation more in the direction of other trading group members in the Large market than in the Small market.

## 4 Concluding discussion

In this paper, we find supportive evidence for the potential of markets to reduce the effects of anchoring on valuations.<sup>11</sup> However, in our study, the market was not needed to correct a potential individual bias due to anchoring. We find no effect of anchoring on reported valuations. This result raises doubt on the robustness of the anchoring effect on people's preferences. Before we discuss differences between our design and other similar studies, we emphasize that our null result is not a consequence of an under-powered study. The observations of the first two sessions were used as a pilot to conduct a power analysis. We conducted a power analysis with an aim to obtain a significant result at the 5% level with 80% power in each of our market treatments separately. The power analysis resulted in an estimated sample size of 148 subjects per market treatment, so 296 subjects in total. To be on the safe side, we aimed for 320. Given that the anchoring hypothesis is based on the total sample (as no treatment has been introduced yet), we have a very high power of 99%.

The existing literature provides a mixed picture of whether anchors affect people's valuations. Studies differ in detail in how they were run, and it is possible that anchoring effects on preferences occur in some circumstances but not in others. When trying to make sense of previous results on anchoring, a complicating factor is that many of these are based on rather small samples which makes it impossible to distinguish between true results, false positives and false negatives. However, even the large studies provide mixed results.

To make sense of previous results, we carried out a limited meta-analysis. In this analysis, we restricted our attention to the 1493 papers that cite Ariely et al. (2003). Among those, we selected the ones that reported an incentivized experiment investigating the effect of anchoring on elicited valuations. We left out results based on hypothetical payments, which includes a very large literature on the effects of anchoring on contingent valuations.<sup>12</sup> Table 3 lists the 19 selected studies on the anchoring of preferences, together with their main features and the reported effect. Each study result is summarized in two ways: (i) as a ratio of valuations of

---

<sup>11</sup> Subjects change their valuations in the direction of the others in their trading group. Our findings corroborate the results of Tufano (2010) and Isoni et al. (2016) who show that markets shape preferences for tasting an unpleasant liquid and preferences for lotteries, respectively. Our results show that markets not only change elicited preferences for unfamiliar goods, i.e. goods where people might not have a clear preference to start with, but also for familiar goods. Our results do not shed light on the question whether the shaping of preferences is a rational process or not. Behavioral conformism may drive the changes in elicited preferences. However, it may also be that preferences for the wine are partly determined by an estimate of the price of the wine in the store, and that people use others' trading decisions to rationally form a better estimate of the retail price.

<sup>12</sup> Some studies combine incentivized treatments and hypothetical treatments. In those cases, our selection only includes results from the incentivized treatments.

High- over Low-anchor group, and (ii) as Hedge's  $g$ , defined as the difference in valuations between High- and Low-anchor group divided by the pooled standard deviation.

The first feature describes the type of good for which a valuation was elicited. *Familiar* goods are ordinary goods that most people now and then consume, like wine, chocolate and books. *Unfamiliar* goods are goods for which people lack daily life experience, like consuming badly tasting liquids and listening to unpleasant sounds. People's preferences may be more affected by anchors when they report their value for an unfamiliar product for which they do not have well-articulated preferences.

The second feature describes the extent to which the anchor may have been perceived as being informative about the price of the good. Some studies use *informative* anchors. One such example is provided by Jung et al. (2016), who investigate the effect of a default anchor on people's donation in a Pay-What-You-Want pricing scheme. Naturally, a default may be perceived as a recommended donation. There are also studies that intended to provide an uninformative anchor, but which may unintentionally have been interpreted as informative by subjects. We categorize these anchors as *questionable*. One such approach is to let a subject's anchor be determined by the last two digits of their social security number (SSN) (e.g., Ariely et al. 2003; Bergman et al. 2010). About one-third of the subjects of Chapman and Johnson (1999) mention that they thought that the SSN anchor was informative. Likewise, Yoon and Fong (2019) and Yoon et al. (2019) use randomly generated uninformative prices, but leave subjects in the dark about the nature of the random number. Their instructions do not exclude the possibility that the random number is somehow correlated to the true price.<sup>13</sup> Studies that use a randomly generated anchor, and clearly communicate the whole procedure to the subjects, are categorized as using an *uninformative* anchor.

The third feature in which studies differ is whether subjects' willingness-to-accept (WTA) or willingness-to-pay (WTP) is elicited. We include this variable because initially there was some support for the idea that anchoring effects are more easily observed for WTP than for WTA (Simonson and Drolet 2004).

Before presenting the results of this concise meta-analysis, we motivate some methodological choices. First, all results need to be weighted appropriately with respect to their precision. Studies are typically weighted by the inverse of the variance of the estimated effect size. We use this approach here and weigh each study according to the variance of Hedge's  $g$ .<sup>14</sup> Second, to test whether the effect size varies across different subgroups, we use the  $Q$  statistic. The  $Q$  statistic is a measure of the weighted variance of the effect sizes and is compared with the variance that

<sup>13</sup> They instructed their subjects about the anchoring in the following way: "First, we will ask whether you would like to buy the item at a particular price. That price will be determined randomly by having you convert the numbers on the card you received into a whole-dollar price."

<sup>14</sup> Hedge's  $g$  is computed as  $g = \frac{m_1 - m_2}{s}$  where  $m_1, m_2$  are the means of the two groups and  $s$  is their pooled standard deviation. The variance of the estimator is given by  $\text{Var}(g) = \frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2)}$ , where  $n_1, n_2$  are the sample sizes of the two groups. The weight of each study is given by  $w = \frac{1}{\text{Var}(g)}$ .

would have been observed if all effect sizes were sampled from a population with the same mean.<sup>15</sup> Third, we use a random-effects model. Given that we are accumulating data from different studies that were carried out in different ways, we believe that the random-effects model is more appropriate than the fixed-effects model.

Figure 3 provides an overview of how these three dimensions affect the effect of anchoring on elicited valuations. We present the results in two separate forest plots, the lower one for studies that use uninformative anchors and the upper one for studies that use informative or questionable anchors. Overall, whether an anchor is uninformative or not has a strong effect on whether anchoring affects elicited valuations or not.<sup>16</sup> With uninformative anchors, we find a precise null-result for the effect of anchoring on elicited valuations. In contrast, with informative or questionable anchors there is a sizable and significant effect of anchoring. The difference in anchoring effects between studies with uninformative anchors and the other studies is significant ( $Q = 27.67, p < 0.001, N = 83$ ).

Within the class of studies that use informative or questionable anchors, we find the following results of the mediating variables on the empirical relevance of anchoring: (i) anchors have a significantly stronger effect for unfamiliar goods for which people do not have a clear initial preference to start with than for familiar goods, whereas (ii) whether WTA or WTP is used to elicit subjects' preferences does not matter; (iii) somewhat surprisingly, we find a significantly stronger effect of questionable anchors than familiar anchors, which supports the view that many of the questionable anchors were actually interpreted as informative by subjects.

Interestingly, in studies that use transparently uninformative anchors, anchoring never has an effect on elicited valuations. So far, whether a familiar or unfamiliar good is used does not matter when the anchor is clearly uninformative. Although these studies are based on relatively many data, there are only few of them, and clearly more studies in this category are welcome.

In the class of studies that use familiar goods, a final interesting comparison is between studies that use clearly uninformative anchors and those that do not. Anchoring effects on elicited valuations are only observed in the latter category, and the difference in the effect is significant ( $Q = 23.21, p < 0.001, N = 52$ ). So, preferences for familiar goods can be anchored, but this requires the use of an anchor that is perceived as informative.

Overall, this meta-analysis yields the following picture: (i) if anchors are informative or perceived to be informative, then (unsurprisingly) anchoring has an effect, and mediating variables play mostly a sensible role, that is, no difference in the effect of anchoring on WTA and in the effect on WTP, while anchors have a stronger effect for valuations of unfamiliar products than familiar products; (ii) in the few

<sup>15</sup> The  $Q$  statistic is computed as  $Q = \sum_{i=1}^k (w_i(ES_i - \bar{ES})^2)$ , where  $k$  is the number of subgroups compared,  $w_i$  is the weight of each study,  $ES_i$  is the effect size of each study and  $\bar{ES}$  is the mean effect size across all studies. Under the null hypothesis that all effect sizes are equal, the  $Q$  statistic follows a  $\chi^2$  distribution with  $(k - 1)$  degrees of freedom. For more details, we refer to (Chap. 8 Card 2015).

<sup>16</sup> As a robustness check, we repeated the analysis without Jung et al. (2016) which has very large weight (due to the total number of participants exceeding 19,000); all conclusions remain qualitatively the same.

**Table 3** Anchoring on valuation studies (ordered by publication year)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Study	Good	Anchor	WTP/WTA	#Results	Sample	Ratio(H/L)	Hedge's <i>g</i>
(a) Informative and questionable anchors (corresponds to Fig. 3a)							
Ariely et al. (2003)	Familiar	[0.2%] Questionable	WTP [2.4%]	1	55	1.71	0.841
Ariely et al. (2003)	Unfamiliar	[15.0%] Informative	WTA [0.9%]	4	61	2.16	1.311
Ariely et al. (2003)	Unfamiliar	[5.7%] Questionable	WTA [3.7%]	1	90	1.62	1.137
Simonson and Drolet (2004)	Familiar	[1.2%] Questionable	WTA [13.2%]	2	139	1.23	0.331
Ariely et al. (2006)	Unfamiliar	[11.7%] Informative	WTP [0.7%]	2	164	1.36	2.905
Ariely et al. (2006)	Unfamiliar	[5.2%] Questionable	WTP [3.4%]	1	81	3.95	1.016
Bergman et al. (2010)	Familiar	[0.5%] Questionable	WTP [5.2%]	1	116	1.43	0.794
Tufano (2010)	Unfamiliar	[9.8%] Informative	WTA [0.6%]	1	134	1.05	0.061
Sugden et al. (2013)	Familiar	[3.9%] Informative	WTA [4.1%]	9	100	1.11	0.165
Sugden et al. (2013)	Familiar	[4.4%] Informative	WTP [4.5%]	9	111	1.09	0.071
Sugden et al. (2013)	Unfamiliar	[13.1%] Informative	WTA [0.8%]	9	20	1.06	0.166
Sugden et al. (2013)	Unfamiliar	[14.6%] Informative	WTP [0.9%]	9	22	1.09	0.086
Koçaş and Demir (2014)	Familiar	[0.1%] Informative	WTP [0.2%]	1	46	5.00	1.766
Maniadis et al. (2014)	Unfamiliar	[8.5%] Informative	WTA [0.5%]	1	116	1.29	0.221
Alevy et al. (2015)	Familiar	[0.8%] Questionable	WTA [8.9%]	1	187	1.06	0.120
Ma et al. (2015)	Unfamiliar	[1.5%] Informative	WTA [0.1%]	1	48	1.34	3.228
Shah et al. (2015)	Familiar	[0.4%] Questionable	WTP [4.1%]	1	95	1.53	0.934
Isoni et al. (2016)	Unfamiliar	[14.9%] Informative	WTA [0.9%]	1	204	1.19	0.186
Jung et al. (2016)	Familiar	[82.2%] Informative	WTP [84.9%]	14	1383	1.13	0.140
Li et al. (2017)	Familiar	[0.4%] Questionable	WTP [4.0%]	1	88	1.21	0.624
Ifcher and Zarghamee (2020)	Familiar	[0.8%] Informative	WTP [0.8%]	1	190	1.31	0.315
Yoon et al. (2019)	Familiar	[1.5%] Questionable	WTP [16.0%]	3	117	1.30	0.615
Yoon and Fong (2019)	Familiar	[3.6%] Questionable	WTP [39.1%]	4	215	1.33	0.581

**Table 3** (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Study	Good	Anchor	WTP/WTA	#Results	Sample	Ratio(H/L)	Hedge's <i>g</i>
(b) Uninformative anchors (corresponds to Fig. 3b)							
Fudenberg et al. (2012)	Familiar	[28.6%] Uninformative	[24.0%] WTA	2	79	1.00	0.001
Fudenberg et al. (2012)	Familiar	[14.1%] Uninformative	[11.8%] WTP	1	78	0.99	0.009
Fudenberg et al. (2012)	Unfamiliar	[100%] Uninformative	[16.4%] WTA	1	108	0.96	0.086
Current study (2020)	Familiar	[57.2%] Uninformative	[47.9%] WTA	1	316	1.05	0.089

Notes to columns (column number in parentheses):

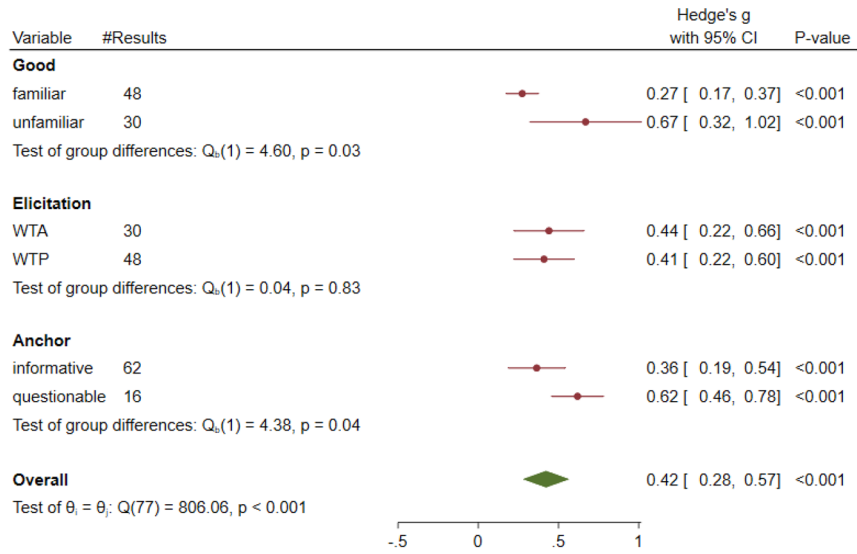
(2–4): The percentages indicate the weight each study receives in the analyses we carry out along each of the three dimensions (see footnote 14 for formulas). The percentages are normalized to add up to 100% within each category. For aggregated results, we present the sum of weights.

(5): #Results display the number of treatments in a paper that share the same features for the variables listed in columns (2), (3) and (4) and that differ in features that are not included in the table.

(6): For aggregated results, we present the average sample size.

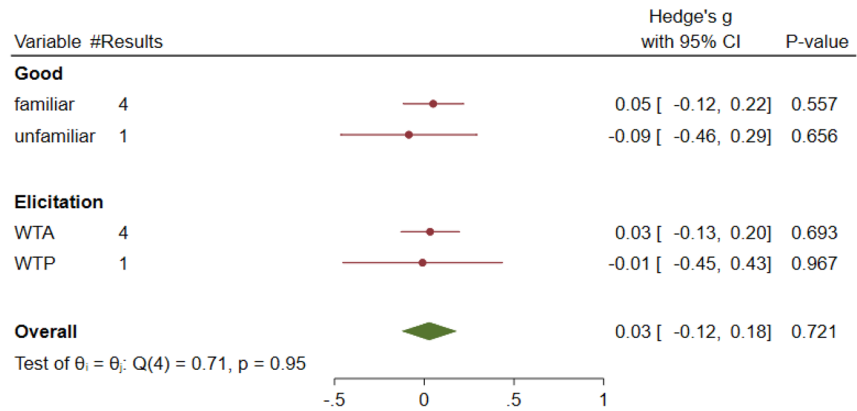
(7): The ratio is computed on the basis of a median split. For aggregated results, we present the sample-weighted average ratio.

(8): For aggregated results, we present the sample-weighted average Hedge's *g*



Random-effects REML model

(a) Informative and questionable anchors



Random-effects REML model

(b) Uninformative anchors

Notes: For each subgroup, the plot includes a dot centered at the mean of the effect size of the corresponding subgroup with lines extended to indicate the confidence intervals. The overall effects are represented by diamonds centered on their estimated values with the width corresponding to the confidence interval length. The height of the diamond is not relevant.

Fig. 3 Forest plots of results by anchor type

studies in which anchors are uninformative, there is a quite precise null-effect, and so far none of the mediating variables plays any role.

In many cases with real-world relevance, evaluations may be made in the presence of seemingly (though not truly) informative anchors. In our view, our study sheds light on what makes anchoring of valuations actually have an impact. Originally,



Tversky and Kahneman (1974) demonstrated the effect of anchors on people's *judgments* with transparently uninformative numbers. On page 1128, they write "subjects were asked to estimate various quantities, stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity, a number between 0 and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given number. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on estimates." Our study suggests that the source of the anchoring of *valuations* may not be that a random uninformative number is imprinted in a subject's mind. Instead, the problem seems to be that people can be tricked into believing that an uninformative piece of information is actually a relevant piece of information. In this sense, the anchoring of valuations is more about people being too gullible when they process information. If the anchoring of preferences only reliably appears when subjects perceive the anchor as informative, then it may be less appropriate to think of the anchoring of preferences as an anchoring bias. Instead, it seems to be driven by a perception bias.

**Acknowledgements** Financial support from the Research Priority Area Behavioral Economics of the University of Amsterdam is gratefully acknowledged. The authors would like to thank participants at the 2019 Economic Science Association European meeting in Dijon, France, and the 2019 CeDEX-CBESS-CREED meeting in Amsterdam, Netherlands, for their valuable comments on the paper. The editor Maria Bigoni and two anonymous referees provided constructive and highly valuable feedback during the review process. We would also like to thank Tina Marjanov for her help in programming the experiment.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alevy, J. E., Landry, C. E., & List, J. A. (2015). Field experiments on the anchoring of economic valuations. *Economic Inquiry*, 53(3), 1522–1538.
- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1), 73–106.
- Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom sawyer and the construction of value. *Journal of Economic Behavior & Organization*, 60(1), 1–10.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9(3), 226–232.
- Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, 107(1), 66–68.
- Bohm, P., Lindén, J., & Sonnegård, J. (1997). Eliciting reservation prices: Becker–degroot–marschak mechanisms vs. markets. *The Economic Journal*, 107(443), 1079–1089.

- Card, N. A. (2015). *Applied meta-analysis for social science research*. New York: Guilford Publications.
- Chapman, G. B., & 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., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9(1), 88–97.
- Fudenberg, D., Levine, D. K., & Maniadiis, Z. (2012). On the robustness of anchoring effects in wtp and wta experiments. *American Economic Journal: Microeconomics*, 4(2), 131–45.
- Ifcher, J., & Zarghamee, H. (2020). Behavioral economic phenomena in decision-making for others. *Journal of Economic Psychology*, 77(1), 102–180.
- Ioannidis, K., Offerman, T., & Sloof, R. (2018). Anchoring bias in markets. American Economic Association's registry for randomized controlled trials. <https://doi.org/10.1257/rct.3402-1.0>.
- Isoni, A., Brooks, P., Loomes, G., & Sugden, R. (2016). Do markets reveal preferences or shape them? *Journal of Economic Behavior & Organization*, 122(1), 1–16.
- Jung, M. H., Perfecto, H., & Nelson, L. D. (2016). Anchoring in payment: evaluating a judgmental heuristic in field experimental settings. *Journal of Marketing Research*, 53(3), 354–368.
- Koçaş, C., & Demir, K. (2014). An empirical investigation of consumers willingness-to-pay and the demand function: The cumulative effect of individual differences in anchored willingness-to-pay responses. *Marketing Letters*, 25(2), 139–152.
- Li, J., Yin, X., Li, D., Liu, X., Wang, G., & Qu, L. (2017). Controlling the anchoring effect through transcranial direct current stimulation (tdcs) to the right dorsolateral prefrontal cortex. *Frontiers in Psychology*, 8(1), 1–9.
- Ma, Q., Li, D., Shen, Q., & Qiu, W. (2015). Anchors as semantic primes in value construction: an eeg study of the anchoring effect. *PloS One*, 10(10), 1–18.
- Maniadiis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: new evidence on anchoring effects. *American Economic Review*, 104(1), 277–90.
- Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological Science*, 26(4), 402–412.
- Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, 31(3), 681–690.
- Sugden, R., Zheng, J., & Zizzo, D. J. (2013). Not all anchors are created equal. *Journal of Economic Psychology*, 39(1), 21–31.
- Tufano, F. (2010). Are 'true' preferences revealed in repeated markets? an experimental demonstration of context-dependent valuations. *Experimental Economics*, 13(1), 1–13.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.
- Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge: Cambridge University Press.
- Yoon, S., & Fong, N. (2019). Uninformative anchors have persistent effects on valuation judgments. *Journal of Consumer Psychology*, 29(3), 391–410.
- Yoon, S., Fong, N. M., & Dimoka, A. (2019). The robustness of anchoring effects on preferential judgments. *Judgment and Decision Making*, 14(4), 470–487.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.