

ORIGINAL RESEARCH REPORT

Do Incidental Environmental Anchors Bias Consumers' Price Estimations?

David R. Shanks*, Pietro Barbieri-Hermitte* and Miguel A. Vadillo†

It is well-established that decision makers bias their estimates of unknown quantities in the direction of a salient numerical anchor. Some standard anchoring paradigms have been shown to yield pervasive biases, such as Tversky and Kahneman's (1974) classic 2-step task which includes a comparative question followed by an estimation question. In contrast there is much less evidence for the claim that incidental environmental anchors can produce assimilative effects on judgments, such as the amount people are willing to pay for a meal being greater at a restaurant called *Studio 97* compared to one called *Studio 17*. Three studies are reported in which the basic incidental environmental anchoring method of Critcher and Gilovich (2008) is employed to measure consumer price estimations. No statistically significant evidence of incidental anchoring was obtained. In contrast, robust standard anchoring effects were found. The results suggest that anchoring is limited to situations which require explicit thinking about the anchor.

Keywords: anchoring; bias; judgment; consumer price estimations

Introduction

How do consumers decide their price estimates and willingness to pay (WTP) for particular goods or services? More generally, how do non-experts form judgments in situations of uncertainty, when the relevant knowledge is, at best, partial?

Attempts to answer these questions have been heavily influenced by the discovery of judgmental biases in which individuals' estimates are systematically skewed away from the true values (Kahneman & Tversky, 2000; Newell, Lagnado, & Shanks, 2015). A prime example is the *anchoring* effect. In a prototypical or *standard* anchoring demonstration reported by Tversky and Kahneman (1974), participants generated what appeared to be an arbitrary number from 0–100 by spinning a roulette wheel. In fact, the wheel was rigged to generate either a low (10) or a high (65) 'anchor' number. Next, they were asked to estimate whether the percentage of African nations in the United Nations was lower or higher than that number. Then they gave their own estimates of this percentage. Remarkably, their final estimates were influenced by the anchor numbers, despite knowing the irrelevance of a roulette wheel's outcome to the true percentage. Those participants who had drawn the low anchor gave lower final estimates than those who had drawn the high anchor.

Judgmental anchoring can be broadly defined as the tendency for decision makers to bias their estimates of unknown quantities in the context or direction of a visually and/or contextually salient value. Typically, anchoring studies adopt a between-subjects 2-step design, where one group of individuals is exposed to a relatively low value of the anchor and the other to a relatively high one. The resulting anchoring effect has been obtained across different contexts and experimental designs (see Chapman & Johnson, 2002; Epley, 2004, for reviews). For example, Jacowitz and Kahneman (1995) found that participants gave greater estimates of the population of Chicago if they had been first asked whether it was less or more than 200,000 than if they had been asked whether it was less or more than 5 million. This effect has been replicated many times including in a large pre-registered multi-lab experiment (Klein et al., 2014), as well as in willingness-to-pay estimates (Simonson & Drolet, 2004; Yoon, Fong, & Dimoka, 2019).

Why does anchoring occur in the 2-step procedure? The idea that it might result from a process of insufficient adjustment was first proposed by Tversky and Kahneman (1974) to explain the findings of the standard paradigm. When making estimations, people will consider whether the value is greater or smaller than the anchor. After this, they *adjust* away from the anchor by producing an estimate. Then, they test this specific value to see if they need to adjust once again – closer to or further from the anchor. Individuals who are satisfied with their answer, and think that they have reached a theoretically plausible estimate, will terminate the adjustment process. As Tversky and Kahneman suggested, this process of

* Division of Psychology and Language Sciences, University College London, UK

† Departamento de Psicología Básica, Universidad Autónoma de Madrid, ES

Corresponding author: David R. Shanks (d.shanks@ucl.ac.uk)

adjustment is typically insufficient, whereby individuals do not adequately adjust away from the anchor. Instead, they remain too close to it.

An alternative hypothesis is that anchoring is the consequence of confirmatory hypothesis-testing. The selective accessibility model (Mussweiler & Strack, 1999; Strack & Mussweiler, 1997) proposes that decision makers judge anchor values as potential answers to be tested and confirmed. As such, anchoring effects arise through a process of semantic priming (but see Harris et al., 2019). Participants are assumed to approach the comparative question by generating a stream of semantic knowledge that is associated with the possibility that the final estimate could actually be equal to the anchor value. For example, in response to the comparative question the participant would ask herself “*Could the population of Chicago really be 5 million?*” and might think about its many skyscrapers and the large number of people who work in them. These activated features then bias the subsequent absolute estimate. When asked in the second step to estimate Chicago’s population, she will rely heavily on whatever knowledge is most accessible, and the activated knowledge (the many skyscrapers) therefore plays a larger role in judgment formation than it would if it had not been activated by the anchor.

Yet another explanation is based on scale distortion, the idea that the anchor changes the psychological interpretation of the response scale (Frederick & Mochon, 2012; but see Harris & Speekenbrink, 2016). This mechanism can be best illustrated by the following example: when asked to judge the weight of a giraffe, participants’ estimates are lowered if they first judge whether it weighs more or less than 20 pounds. Frederick and Mochon (2012) proposed that this effect arises because the small weight in the comparative question makes pounds seem like larger units, and hence a lower estimate is needed to accurately capture the giraffe’s weight. The individual’s sense of the giraffe’s weight is unchanged by the anchor; what changes is the mapping of that sense onto the judgment scale.

Decision makers will infrequently be faced with the somewhat artificial, two-step process of the standard anchoring paradigm when making decisions (though see Jung, Perfecto, & Nelson, 2016). While the 2-step procedure described above is a standard tool in the laboratory, anchoring effects also occur in more naturalistic situations, including ones in which the anchor is entirely non-numerical, and hence when no number processing is involved at all. For example, LeBoeuf and Shafir (2006) asked participants in one condition to add pennies to an empty cup, while those in another condition removed pennies from a cup which initially weighed 12 ounces, until the cup weighed the same as another cup they had held and subjectively weighed earlier and which actually weighed 6 ounces. The starting weight of the cup acted as an anchor such that final cup weights were larger in the group adjusting downwards from a high anchor (12 ounces) than in the group adjusting upwards from an empty cup (0 ounces).

In another example, which they termed *basic* anchoring, Wilson, Houston, Etling, and Brekke (1996) asked one group of participants to copy a series of high numbers across one page, while another group were asked to do the same across 5 pages. After this, they were all asked how many fellow students they thought would contract cancer over the following forty years. The results showed that those who had copied numbers across more pages gave higher estimates of cancer contraction, compared to those who had copied numbers on fewer pages. Brewer and Chapman (2002), however, found that this effect was fragile and disappeared following trivial changes in methodology.

Incidental environmental anchoring

In the present research we focus on the claim that seemingly irrelevant values encountered in specific scenarios can bias decision making and produce assimilative effects on judgments (Critcher & Gilovich, 2008; Dogerlioglu-Demir & Koçaş, 2015; Koçaş & Dogerlioglu-Demir, 2020; Nunes & Boatwright, 2004) – so called ‘incidental’ or ‘deliberation-free’ (Jung et al., 2016) anchoring. Surprisingly, although garnering many citations, studies exploring this potentially-important effect have been few in number.

Critcher and Gilovich (2008) examined this possibility by asking participants to make judgments about scenarios that were accompanied by photographs incorporating incidental anchors. In their first study a fictitious college linebacker, Stan Fischer, was described alongside a photograph of him wearing a jersey with either the number 54 (low anchor) or 94 (high anchor). Despite the fact that participants were not required to make any explicit judgment about the jersey number (as they would in a conventional anchoring task) – and indeed may have barely registered it – participants nevertheless judged Fischer more likely to register a sack in the conference playoff game in the high than in the low anchor condition. Unlike other anchoring procedures (such as the 2-step one), participants are not required to explicitly think about or reflect on the anchor in the incidental anchoring procedure – hence its name.

In their second and third studies, Critcher and Gilovich (2008) reported similar effects in different domains, including in marketing and willingness-to-pay contexts. Participants in Study 2 gave higher estimates of the percentage of sales of a new mobile phone in the United States versus Europe when the phone was described (with an accompanying photograph) as a *Sony Ericsson P97* than as a *Sony Ericsson P17*. In Study 3 participants were shown a photograph of a restaurant and asked how much they would be willing to pay for a meal at this restaurant. Willingness-to-pay amounts were larger when the restaurant was called *Studio 97* than when it was called *Studio 17*.

Some mediators of these incidental anchoring effects were explored by Dogerlioglu-Demir and Koçaş (2015; Koçaş & Dogerlioglu-Demir, 2020). Study 1 of Dogerlioglu-Demir and Koçaş (2015) replicated the

effect of a number in the name of a restaurant on WTP estimates: Participants expressed higher estimates of how much they would be willing to spend on a meal when shown a picture of *Studio 97* than when shown a picture of *Studio 17*. However the effect was eliminated when the question referred to willingness to pay for 'a hamburger meal' rather than simply 'a meal'. Dogerlioglu-Demir and Koçaş (2015) hypothesized that products differ in the extent to which people maintain internal reference prices (IRPs) and that when the IRP is strong (as in the case of a hamburger, for which most people have strong price expectations), the influence of an incidental anchor will be attenuated. An anchoring effect was also obtained when the incidental number appeared in the restaurant's address (*Studio A, 17th Street*) instead of its name.¹

Incidental anchoring merits further study because the phenomenon has major implications (e.g., in marketing) and yet the available evidence is weak. For instance, the effects documented by Critcher and Gilovich (2008) were remarkably fragile. As Matthews (2011) noted, from a Bayesian statistical perspective (Dienes, 2011) the evidence they reported provides hardly any more support for the experimental hypothesis (anchoring) than for the null hypothesis (no anchoring). Reanalysis of the data from Dogerlioglu-Demir and Koçaş' (2015) Study 1 finds that the statistically significant effect they reported is no longer significant even by a 1-tailed test, $t(55) = 1.46$, $p = .08$, when two datapoints identified by boxplots as outliers are removed.

In the same vein, these studies seem to be relatively underpowered to detect the kind of effects that one would expect to find in typical psychology experiments, which usually yield medium-to-small effect sizes, in the range of 0.40 to 0.50 in Cohen's d units (Bakker, van Dijk, & Wicherts, 2012). As we report later, the meta-analytic average effect of the three experiments that have explored the impact of incidental anchoring on WTP is $d = 0.49$. The sample recruited by Critcher and Gilovich (2008, Study 3) provides excellent statistical power (i.e., .92) to detect an effect of this size. In contrast, Studies 1 and 1b from Dogerlioglu-Demir and Koçaş (2015) achieve power of only .46 and .25, respectively. Given these values, the probability of obtaining a significant result in all three studies is just .11. Although this probability does not reach conventional significance, it is low enough to suggest that the published record on this topic might be overoptimistic, perhaps due to the selective publication of studies (or analyses) with significant results (Francis, 2012).

More significantly, a recent multi-lab replication project (Many Labs 2: Klein et al., 2018) failed to replicate the findings of Critcher and Gilovich's (2008) Study 2. As part of a large battery of tests presented in a random order, participants saw an updated version of Critcher and Gilovich's materials comprising a picture of a smartphone with the model number *Sony Ericsson P97/P17* on the phone's display and some background text describing the smartphone before estimating what percentage of sales would be in the United States, as opposed to

Western Europe. Interestingly, this replication failure was anticipated in advance by the contributors to the Many Labs 2 project, who participated in a prediction market and stated their beliefs about the reproducibility of 24 effects prior to data collection (Forsell et al., 2019). Both the prediction market and the stated beliefs correlated highly with actual reproducibility across these 24 effects (for instance, $r = 0.76$ for the prediction market), and both of them judged the reproducibility of Critcher and Gilovich's (2008) Study 2 as being less than 50%.

The failure to obtain an incidental anchoring effect in the Klein et al. (2018) study places a substantial doubt over the reproducibility of the original demonstrations. However it does not directly address the question at issue in the present work, namely whether incidental anchors can bias price or WTP judgments. These plainly are more frequent and of greater practical significance than judgments about the percentage of sales of a product in one country compared to another. It is known that different judgments are differentially sensitive to standard anchoring effects. For instance, Simonson and Drolet (2004) found that WTP but not willingness to accept judgments were influenced by an anchor. Hence the lack of an incidental anchoring effect on percentage sales estimates provides little information about whether or not such anchors can affect price or WTP judgments. More importantly, the incidental anchoring effect on WTP reported by Critcher and Gilovich (2008, Study 3) and Dogerlioglu-Demir and Koçaş (2015) was substantially larger than the effect on percentage sales ($d = 0.49$ vs. 0.30), as well as being documented in 3 independent studies compared to one. Hence Klein et al.'s failure to obtain an incidental anchoring effect on percentage sales estimates provides little reason to question the positive effects on WTP reported by Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koçaş (2015). There remains a pressing need for further studies of incidental anchoring, particularly on price and WTP estimates (for discussion of methods for measuring WTP, see Wertebroch & Skiera, 2002). The present research aims to fill this gap.

For all experiments reported in this article, we report how we determined our sample size, all data exclusions, all manipulations, and all measures. All data and materials are publicly available via the Open Science Framework (OSF) at osf.io/8ynwu. All studies reported here were approved by the University College London (UCL) Research Ethics Committee.

Study 1

Study 1 aims to replicate the design of Study 3 by Critcher and Gilovich (2008) and Study 1 by Dogerlioglu-Demir and Koçaş (2015), asking participants to report their willingness to pay for a meal at a restaurant described as either *Studio 97* or *Studio 17*. In addition to the restaurant item, participants gave price estimations for 6 other goods and services, including technology items similar to those used in Critcher and Gilovich's Study 2. Based on their findings, we expect all high anchor items to produce higher price estimations than the low anchor items.

Method

Participants and design

We employed a two-condition (high vs low anchor) between-subjects design to test the effect of different anchors on participants' price estimations for 7 different goods and services. Participants were recruited via Amazon Mechanical Turk (Mturk) and were paid \$0.50. We planned all studies' sample sizes prior to data collection. The effect obtained in previous studies was medium-sized (Cohen's $d \approx 0.5$).² On the basis of a power analysis, we aimed to recruit 70 participants per group in order to achieve high power $1-\beta = 0.90$ to detect a medium-sized effect at $\alpha = .05$, 1-tailed. A total of 144 participants (97 males and 47 females) with a minimum qualification of 1,000 previously 'accepted' human intelligence tasks (HITs) completed the survey. All participants were from the United States and reported their price estimations in US dollars, except for 1 participant who gave price estimations in rupees (which were converted to dollars for analysis). The median time to complete the questionnaire was approximately 3 min.

Materials

The survey consisted of 7 images of selected consumer products (see Supplemental Materials at osf.io/8ynwu). Two versions of each were created, one containing a low anchor and the other a high anchor. Only the numeric value associated with each item was manipulated. The exact wording of each item is provided in Table A1 (Additional File).

The first image (Restaurant) was a picture of a restaurant, which either contained the name *Studio 17* or *Studio 97* at the top right-hand side of its exterior. The second one (Hotel) was a picture of a mid-range hotel room, with either *No. 44 Cranley Place* or *No. 404 Cranley Place* displayed at the top right-hand corner; the third image (Membership) consisted of a picture of the Facebook logo including either *2.0* or *2020* in the same font as the original logo just below it; the fourth image (VR glasses) was of a prototype of virtual reality glasses containing *Fujitsu 400X* or *Fujitsu 4000X* at the bottom left-hand corner; the fifth image (iPad) was of a prototype of a see-through iPad, labelled either the *iPad 66* or *iPad606* at its top right-hand corner; the sixth stimulus (Bar) was a picture of a bar named either *Tap 42* or *Tap 92*; and finally, the seventh stimulus (Holiday) was a picture of an exotic beach resort, including the Club Med logo with either *102* or *10210* added next to it in the same font at the top left-hand corner of the image.

The images were divided into two sets in order to counterbalance the high- and low-anchor versions. Set A comprised: Restaurant/low, Hotel/high, Membership/low, VR glasses/high, iPad/low, Bar/high, Holiday/low; Set B comprised: Restaurant/high, Hotel/low, Membership/high, VR glasses/low, iPad/high, Bar/low, Holiday/high.

Procedure

The software decided at random whether to present each participant with the images from Set A or Set B. Whichever set was chosen, the presentation order of the items was

always the same. The first item was always a picture of the restaurant labelled *Studio 17/Studio 97* (Restaurant), followed by the Hotel, Membership, VR glasses, iPad, Bar, and finally the Holiday question. Thus each participant saw both high- and low-anchor items, but for each item the key contrast is the between-subjects comparison of ratings.

Beneath 5 of the images brief text (see Table A1) related to that specific consumer good or service was provided. For example, just below the iPad item, participants read: "*Apple are planning to release transparent technology on all their gadgets in the near future. How much would you be willing to pay for this product?*"; and underneath the Club Med item, they read: "*Club Med is planning to organise reduced-cost holiday packages to exotic destinations. How much would you be willing to pay for this package?*" This text provided context for the participants to judge how much they would be willing to pay for the good or service. However, for the Restaurant and Bar items no text was provided. Instead, participants were simply asked: "*How much would you be willing to pay for your own meal at this restaurant?*"

After participants typed in a number as their estimate for all 7 items, they were asked to describe what they thought the purpose of the survey was. This was probed so we could exclude any participants who expressed awareness of the anchoring manipulation. Awareness was minimal in both this study and Studies 2 and 3 with most participants either expressing no idea of the purpose, or stating that it was to gather information on people's estimates about everyday purchases and events.

Results

Estimates entered as text were converted to numerical format (e.g., "2 million" to 2,000,000). Outlier price estimates, defined as observations falling more than 1.5 times the interquartile range (IQR) from the extreme of the IQR, were identified via the 'boxplot' function in R (R Core Team, 2018). In total 5.56% of observations were excluded. Independent-samples 1-tailed t -tests were conducted to compare the average price estimations for high and low anchor items. For each item, the independent variable was the anchor value (low vs. high). Descriptive and inferential statistics are reported in **Table 1**.

The Restaurant high anchor group gave price estimations that were numerically higher than those of the low anchor group but the difference fell far short of statistical significance ($p = .26$), and a Bayesian analysis in fact indicated somewhat stronger support for the null than the experimental (H1: high anchor estimates > low anchor estimates) hypothesis, $BF_{+0} = 0.33$. (All Bayes factors reported here adopt a Cauchy prior distribution with scale r on effect size equal to 0.707).

For the VR glasses, iPad, and Holiday items anchoring effects in the predicted direction were also observed. In only one of these cases however did the effect provide statistical support for the experimental hypothesis: For the iPad item $BF_{+0} = 2.66$, meaning that the experimental hypothesis is about 3 times more likely than the null hypothesis given the data. Although the effect is statistically significant ($p = .02$), it would not survive a

Table 1: Results of Experiment 1.

| Item-pair | Anchor condition | | | | | | Cohen's <i>d</i> | 95% CI | <i>t</i> | df | <i>BF</i> ₁₀ |
|-------------------|------------------|-----------|----------|----------|-----------|----------|------------------|--------------|----------|-----|-------------------------|
| | High | | | Low | | | | | | | |
| | <i>M</i> | <i>SD</i> | <i>n</i> | <i>M</i> | <i>SD</i> | <i>n</i> | | | | | |
| <i>Restaurant</i> | 16.07 | 6.40 | 67 | 15.36 | 6.35 | 69 | 0.11 | -0.23, 0.45 | 0.65 | 134 | 0.33 |
| <i>Hotel</i> | 118.75 | 45.95 | 69 | 118.77 | 45.97 | 71 | 0.00 | -0.33, 0.33 | 0.00 | 138 | 0.18 |
| <i>Membership</i> | 2.77 | 3.75 | 66 | 4.43 | 5.30 | 65 | -0.36 | -0.71, -0.02 | -2.07 | 129 | 0.06 |
| <i>VR glasses</i> | 110.24 | 97.27 | 68 | 103.78 | 81.85 | 69 | 0.07 | -0.26, 0.41 | 0.42 | 135 | 0.26 |
| <i>iPad</i> | 317.96 | 229.58 | 73 | 242.09 | 192.43 | 67 | 0.36 | 0.02, 0.69 | 2.11 | 138 | 2.66 |
| <i>Bar</i> | 18.63 | 8.37 | 67 | 18.97 | 7.86 | 68 | -0.04 | -0.38, 0.30 | -0.25 | 133 | 0.15 |
| <i>Holiday</i> | 757.77 | 493.08 | 67 | 713.96 | 339.91 | 66 | 0.10 | -0.24, 0.44 | 0.60 | 131 | 0.31 |

correction for multiple comparisons. Moreover in Study 2 we will see that the equivalent effect is not replicated. For the VR glasses and Holiday items the evidence actually supported the null more than the experimental hypothesis (i.e., $BF_{10} < 1$).

For the Membership and Bar items price estimates were numerically greater for the low than high anchor versions. Overall for 6/7 items the Bayes factor analysis favoured the null hypothesis.

Discussion

Study 1 suggests that numbers encountered in the environment do not exert a strong influence on judgments of value. We were unable to find any statistically significant difference in price estimations between *Studio 17* and *Studio 97*. The estimations of participants who were exposed to a lower anchor compared to those who were exposed to a higher one were not biased by or assimilated to their respective anchors. This is particularly interesting given that we employed a very similar procedure to the one used by both Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koças (2015). Before drawing firm conclusions, we sought to replicate these null effects. Study 2 was similar to Study 1 but adopted question wording closer to that of Critcher and Gilovich.

Study 2

In Critcher and Gilovich's (2008) Study 3, the restaurant item was always followed by a question which repeated its name, either *Studio 17* and *Studio 97*, in the question. Thus in Study 2 the wording was modified compared to that used in Study 1 (see Table A1). Below the restaurant picture, participants were asked: "Studio ___ is a new restaurant. The above picture was taken from a magazine advertisement. Imagine that you are going to dine at this restaurant. Estimate how much your bill would amount to (not your entire party's, just yours)". Note that this is a price rather than WTP estimation. As detailed in Table A1, in this study we employed both types of estimates across the different items.

Another modification is that we added two standard anchoring trials after all the incidental ones. For these items, of course, we strongly predict that participants given

high anchors will generate higher absolute estimates than ones given low anchor values.

Method

Participants and design

On the basis of Study 1 we sought to recruit an adequate sample to detect a smaller incidental anchoring effect (Cohen's $d \approx 0.3$). A power analysis indicated that 139 participants per group would yield acceptable power ($1-\beta = 0.80$) to detect such an effect at $\alpha = .05$, 1-tailed. Again, we recruited US participants through Mturk, all of whom had previously achieved a minimum of 97% on their approval rate, and a previous completion of at least 1,000 surveys on Mturk. A total of 273 individuals (153 males) completed the study. The majority reported their price estimations in US dollars, except for 25 who gave estimations in rupees and two in Euros (these were converted to dollars for analysis). The median time to complete the questionnaire was approximately 5 min.

Materials

The 14 images (7 pairs) used in Study 1 were used once again in the second study with the same high and low anchor values. However, in an attempt to replicate the image and text design of Critcher and Gilovich (2008) even more closely, the wording of each question was modified (see Table A1) by explicitly repeating the anchor value in the question. In addition, participants saw two additional items after they had completed the incidental trials. These presented standard anchoring questions using simple, two-part questions, each one appearing in either a high or a low anchor version. One item (Population) consisted of an estimation of the population of Chicago, and read: "Is the population of Chicago (IL) greater or less than 1,500,000/5,000,000?", after which participants either chose "Greater than" or "Less than", and then they answered the question: "What is the population of Chicago (IL)?" The second item (Length) had a similar format but asked about the length of the Amazon River, with anchors of 3,000 miles (low) and 6,000 miles (high).

The questions were again divided into two sets in order to counterbalance the high- and low-anchor versions. Set A comprised: Restaurant/low, Hotel/high, Membership/low, VR glasses/high, iPad/low, Bar/high,

Holiday/low, Population/high, Length/low. Set B comprised: Restaurant/high, Hotel/low, Membership/high, VR glasses/low, iPad/high, Bar/low, Holiday/high, Population/low, Length/high.

Procedure

The flow of trials was similar to that in Study 1. Once again, this involved an initial randomisation of the Restaurant question, and then a counterbalancing of the high and low anchor versions of each question. Participants answered in a text box underneath the script of each question. The two final standard anchoring trials were added to the survey flow after all the incidental trials. These items enabled a replication of the classic anchoring procedure designed by Tversky and Kahneman (1974).

Results

Estimates entered as text were again converted to numerical format. Two estimates were ambiguous (e.g., “8,00,000”) and hence were treated as missing values. Outlier estimates (7.37% of observations) were identified in the same way as in Study 1. Independent-samples *t*-tests were conducted to compare the average estimations for high and low anchor items. Descriptive and inferential statistics are reported in **Table 2**.

There was no significant anchoring effect for the Restaurant (*Studio 97* vs. *Studio 17*) picture with almost identical estimates in the high and low anchor groups, and the Bayesian analysis indicated fairly strong support for the null over the experimental (high anchor estimates > low anchor estimates) hypothesis, $BF_{+0} = 0.15$.

For the Hotel, VR glasses, and Bar items price estimates were numerically greater for the high than low anchor versions. In none of these cases however did the effect provide clear statistical support for the experimental hypothesis. For the Membership, iPad, and Holiday items anchoring effects in the nonpredicted direction were observed. For the iPad item $BF_{+0} = 0.05$, meaning that

the null hypothesis is over 10 times more likely than the experimental hypothesis given the data, and strongly implying that the modest effect observed in Study 1 for this item was a sampling artifact. For all 7 incidental anchoring items the Bayes factor analysis favoured the null hypothesis.

Standard anchoring questions

A further two independent samples *t*-tests were conducted to compare the average estimates for high and low anchor stimuli on each of the two standard anchoring items. Once again, the independent variable was the anchor value, with two levels, high and low. The dependent variable was the average estimate. The results are reported in **Table 2**.

In contrast to the incidental anchoring results, for these standard items robust anchoring effects were obtained ($p < .001$ in each case). As expected, the group who were given the high anchor (5,000,000) for the population of Chicago gave average estimations that were much larger (by nearly 30%) than the group who were given the low anchor (1,500,000). Likewise, the group given a high anchor (6,000 miles) for the length of the Amazon River made average length estimates that were much longer (again by about 30%) than the group given the low anchor (3,000 miles). The Bayesian analysis shows that in each case the evidence provides strong support for the experimental hypothesis (i.e., an anchoring effect).

Discussion

Once again, Study 2 suggests that numbers encountered incidentally in the environment do not have any significant effect on judgments of value. Participants did not estimate that a meal at *Studio 97* would be more expensive than one at *Studio 17*, nor were they willing to pay more for any of the other goods or services, despite our efforts to closely replicate the procedure adopted by both Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koças (2015).

Table 2: Results of Experiment 2.

| Item-pair | Anchor condition | | | | | | Cohen's <i>d</i> | 95% CI | <i>t</i> | df | BF_{+0} |
|-------------------------------------|------------------|-----------|----------|----------|-----------|----------|------------------|--------------|----------|-----|--------------------|
| | High | | | Low | | | | | | | |
| | <i>M</i> | <i>SD</i> | <i>n</i> | <i>M</i> | <i>SD</i> | <i>n</i> | | | | | |
| Incidental | | | | | | | | | | | |
| <i>Restaurant</i> | 21.54 | 11.58 | 127 | 21.43 | 10.26 | 125 | 0.01 | -0.24, 0.26 | 0.07 | 250 | 0.15 |
| <i>Hotel</i> | 109.00 | 54.47 | 127 | 103.94 | 57.32 | 126 | 0.09 | -0.16, 0.34 | 0.72 | 251 | 0.27 |
| <i>Membership</i> | 10.33 | 9.01 | 115 | 14.21 | 14.10 | 131 | -0.32 | -0.58, -0.07 | -2.53 | 244 | 0.04 |
| <i>VR glasses</i> | 162.52 | 108.03 | 123 | 152.88 | 111.36 | 131 | 0.09 | -0.16, 0.33 | 0.70 | 252 | 0.26 |
| <i>iPad</i> | 380.18 | 189.79 | 128 | 438.53 | 250.19 | 129 | -0.26 | -0.51, -0.02 | -2.11 | 255 | 0.05 |
| <i>Bar</i> | 26.36 | 11.14 | 124 | 24.83 | 11.92 | 127 | 0.13 | -0.12, 0.38 | 1.05 | 249 | 0.39 |
| <i>Holiday</i> | 817.31 | 545.90 | 133 | 868.08 | 592.48 | 129 | -0.09 | -0.33, 0.15 | -0.72 | 260 | 0.08 |
| Standard | | | | | | | | | | | |
| <i>Population</i> ($\times 10^6$) | 3.41 | 1.86 | 123 | 2.54 | 0.76 | 113 | 0.60 | 0.34, 0.86 | 4.61 | 234 | 4782 |
| <i>Length</i> ($\times 10^3$) | 5.16 | 2.08 | 133 | 3.85 | 1.26 | 130 | 0.76 | 0.51, 1.01 | 6.15 | 261 | 6.72×10^6 |

As expected, both of the standard anchoring items produced large differences in judgments between items with low relative to high comparative anchor values. Plainly, under conditions in which standard anchoring is robust, incidental anchoring is negligible.

Study 3

Study 3 had several aims. The first was to obtain data from a large sample with a fully preregistered method and procedure. The second was to eliminate some additional small but potentially important differences between the procedure used in Studies 1 and 2 and that employed by Critcher and Gilovich (2008). Chief amongst these is that whereas the Restaurant question in Studies 1 and 2 was illustrated by a picture of the exterior of Studio 17/97 with its name (and hence the anchor) plainly visible, Critcher and Gilovich used an interior picture and only stated the restaurant's name in the accompanying question (e.g., "Studio 17 is a new restaurant..."). Thus the anchor was extrinsic to the image in Critcher and Gilovich's Study 3, but intrinsic in the present Studies 1 and 2. It seems unlikely that such a minor change would alter the results, especially bearing in mind that Critcher and Gilovich reported incidental anchoring effects with intrinsic images (e.g., of a footballer wearing a jersey with either 54 or 94 on it) in their Studies 1 and 2, but we nevertheless sought to eliminate this difference.

In Study 3 we included 2 pairs of standard anchoring items. One pair employed the same presentation method as in Study 2 in which a comparative question was followed by an estimation question. Unlike in Study 2, we presented the 2 questions on separate pages so that any anchoring effect would have to be carried over from memory of the comparative question. In the other pair the presentation format was altered to make the items more similar to the incidental anchoring format. Specifically, participants were asked just a single estimation question (e.g., "The population of Chicago is [more than 200,000/less than 5,000,000]. What do you think the population of Chicago is?"). This is the format used by Klein et al. (2014) in their successful multi-lab replication of anchoring.

Method

The experiment was preregistered at osf.io/e48bu.

Participants and design

A power analysis indicated that 429 participants per group would yield high power ($1-\beta = 0.90$) to detect a small (Cohen's $d = 0.2$) incidental anchoring effect at $\alpha = .05$, 1-tailed. Again, we recruited US participants through Mturk, all of whom had previously achieved a minimum of 95% on their approval rate. A total of 874 individuals (434 males; 3 withheld their gender; mean age 36.7 years) completed the study. The median time to complete the questionnaire was approximately 3 min.

Materials

There was only one incidental anchoring question in this study, the Restaurant item. Whereas the restaurant image employed in Studies 1 and 2 was of its exterior

and named the restaurant (and hence the anchor), Critcher and Gilovich (2008) as well as Dogerlioglu-Demir and Koçaş (2015) used interior images which did not include the anchor. We therefore also used an interior image in this study (see Supplemental Materials at osf.io/8ynwu). This was chosen to be quite similar visually to that of Critcher and Gilovich, although with a more contemporary atmosphere and higher image quality. The restaurant name (and hence anchor) was not included in the image.

In Study 2 the question wording stated that "Studio ___ is a new restaurant". Here the wording was slightly changed to "Studio ___ is a new restaurant set to open in Butler, Pennsylvania", thus making it identical to that used by Critcher and Gilovich. Participants estimated how much they would be willing to pay for a dinner at this restaurant (see Table A1). All payments were explicitly in dollars.

After the Restaurant item, participants responded to four standard anchoring questions (see Table A1), two of which (Birthrate and Year) used two-part questions (e.g., "Was the World Wide Web invented before or after ___? [Before/After]. In what year was the World Wide Web invented?") and two of which (Population and Distance) presented just a single question ("The distance from San Francisco to New York City is longer/shorter than ___ miles. How far do you think it is?"). Each appeared in either a high or a low anchor version. Note that the low anchor for the Population question was reduced compared to Study 2. In contrast to the previous studies, for all items except Birthrate (where participants typed their response), estimates were entered via a slider.

Procedure

The questions were again divided into two sets in order to counterbalance the high- and low-anchor versions. Set A comprised: Restaurant/low, Population/high, Birthrate/low, Distance/high, Year/low. Set B comprised: Restaurant/high, Population/low, Birthrate/high, Distance/low, Year/high.

The flow of trials was similar to that in Study 2. Once again, this involved an initial randomisation of the Restaurant question, and then a counterbalancing of high and low anchor questions, presented in a fixed order.

Results

There were no deviations from the preregistered plan and no missing values. Outlier estimates (4.62% of observations) were identified in the same way as in Study 1. Independent-samples t -tests were conducted to compare the average estimations for high and low anchor items. Descriptive and inferential statistics are reported in **Table 3**.

There was no significant anchoring effect for the Restaurant (*Studio 97* vs. *Studio 17*) question with almost identical mean estimates in the high and low anchor groups, and the Bayesian analysis indicated strong support for the null over the experimental (high anchor estimates > low anchor estimates) hypothesis, $BF_{+0} = 0.07$. This conclusion remains ($BF_{+0} = 0.16$) in an exploratory

Table 3: Results of Experiment 3.

| Item-pair | Anchor condition | | | | | | Cohen's <i>d</i> | 95% CI | <i>t</i> | df | <i>BF</i> ₊₀ |
|---------------------------------------|------------------|-----------|----------|----------|-----------|----------|------------------|-------------|----------|-----|-------------------------|
| | High | | | Low | | | | | | | |
| | <i>M</i> | <i>SD</i> | <i>n</i> | <i>M</i> | <i>SD</i> | <i>n</i> | | | | | |
| Incidental | | | | | | | | | | | |
| <i>Restaurant</i> | 30.18 | 14.28 | 406 | 30.38 | 14.20 | 422 | -0.01 | -0.15, 0.12 | -0.20 | 826 | 0.07 |
| Standard | | | | | | | | | | | |
| <i>Population</i> (×10 ⁶) | 3.21 | 1.05 | 437 | 2.05 | 1.10 | 435 | 1.08 | 0.94, 1.23 | 15.99 | 870 | 2.50 × 10 ⁴⁷ |
| <i>Birthrate</i> (×10 ³) | 43.59 | 36.56 | 392 | 1.49 | 1.63 | 365 | 1.60 | 1.43, 1.76 | 21.98 | 755 | 4.18 × 10 ⁷⁹ |
| <i>Distance</i> (×10 ³) | 3.82 | 1.00 | 437 | 3.15 | 0.86 | 417 | 0.72 | 0.58, 0.86 | 10.49 | 852 | 2.35 × 10 ²¹ |
| <i>Year</i> | 1988.44 | 5.51 | 418 | 1984.12 | 7.62 | 439 | 0.65 | 0.51, 0.78 | 9.47 | 855 | 2.84 × 10 ¹⁷ |

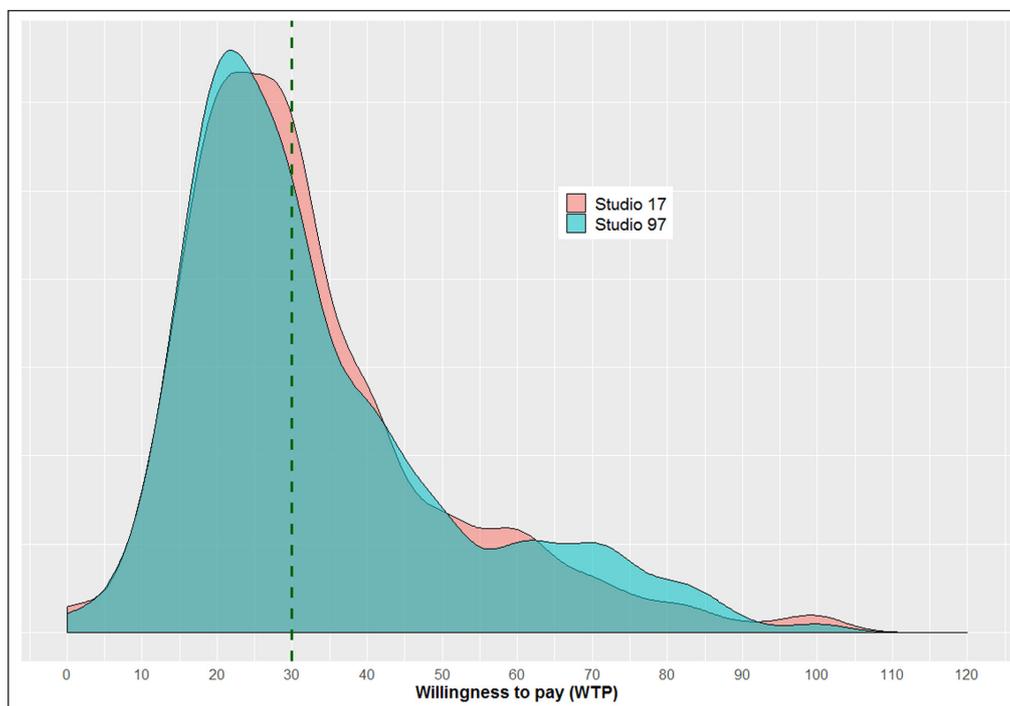


Figure 1: Density plots of Willingness to pay (WTP) estimates in dollars in the two Restaurant conditions of Study 3. All data (including outliers) are plotted. The dotted line is the median estimate (identical in the Studio 17 and Studio 97 conditions).

analysis which drops the preregistered outlier exclusion rule and instead includes all of the data. **Figure 1** presents density plots, with outliers included, confirming the very close similarity in the distributions and medians of WTP estimates in the two conditions.

It is worth noting that while the price estimates made by participants in Studies 1 and 2 (about \$16 and \$21, respectively) are slightly lower than those made by participants in the Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koçaş (2015) studies (about \$30), in Study 3 the mean estimates were higher (about \$30) and close to those reported in the earlier studies. Although the difference could be due to other factors (e.g., the participant samples), it is consistent with a better match of the relatively up-market interior image we presented in this study to those used by Critcher and Gilovich and Dogerlioglu-Demir and Koçaş.

Standard anchoring questions

In contrast to the incidental anchoring results, for these standard items robust anchoring effects were obtained with large *BF*₊₀ values. Anchoring did not differ in any obvious way between the items (Birthrate and Year) which used two-part questions and those (Population and Distance) which presented just a single question. Across these four items, the largest effect (Cohen's *d* = 1.58) came from one of the two-part items (Birthrate) but the smallest effect (*d* = 0.65) came from the other one (Year).

Meta-Analysis

The present results supplement previous ones and permit us to aggregate all available relevant data and maximize power for the purpose of determining what the weight of evidence reveals about incidental anchoring of price/willingness-to-pay estimates.

In total there are 6 studies on these estimates using variants of the methods employed here. These are Critcher and Gilovich's (2008) Study 3, Studies 1 and 1b by Dogerlioglu-Demir and Koças (2015), and the present Studies 1–3. All of these studies presented an image of a restaurant, though of course Studies 1–2 also included other items. Note that we exclude the field studies of Nunes and Boatwright (2004) which use very different methods (we comment further on these in the General Discussion).

We included the effect sizes as reported in Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koças (2015) even though they are inflated by outliers. The effect reported by Critcher and Gilovich (2008, Study 3) reduces

from Cohen's $d = 0.40$ to 0.27 when 5 outliers (identified by boxplots) in the Studio 97 condition are excluded, while the Dogerlioglu-Demir and Koças (2015) effect reduces from Cohen's $d = 0.54$ to 0.39 when 2 outliers in the same condition are excluded.

In order to control for the fact that effect sizes for different items in the present Studies 1–3 are statistically dependent on each other, we ran a multi-level random-effects meta-analysis using the 'rma.mv' function in the R 'metafor' package (Viechtbauer, 2010), adding a random intercept at the study level. The forest plot is shown in **Figure 2**. Across all studies and effects, the mean effect size is Cohen's $d = 0.41$, 95% CI [0.12, 0.69], based on a total of 1574 participants.



Figure 2: Forest plot. The rows denote the effect sizes (ES; Cohen's d) and lower and upper 95% confidence intervals of anchoring effects on value estimates, grouped separately for the original studies, the incidental conditions of Studies 1–3, and the standard anchoring conditions of Studies 2 and 3. The final rows report the meta-analytic effects. C & G: Critcher & Gilovich; D-D & K: Dogerlioglu-Demir & Koças.

The meta-analysis revealed a substantial and statistically significant amount of heterogeneity, $I^2 = 87.92\%$, $Q(23) = 513.23$, $p < .001$. To assess whether this heterogeneity could be accounted for by anchoring type (incidental vs. standard), we conducted a metaregression. This found that anchoring type strongly moderates the results: Effect sizes were much larger for standard ($d=0.84$, 95% CI [0.57, 1.11]) than for incidental ($d = 0.19$, 95% CI [-0.04, 0.42]) anchoring, $Q_M(1) = 202.06$, $p < .001$. When including only the Restaurant studies, the effect is small and the confidence interval extends to zero, $d=0.22$, 95% CI [0.00, 0.45], $z = 1.92$, $p = .055$. This overall weak effect, however, must be interpreted with considerable caution because the meta-analysis also reveals a strong small-study effect. The funnel plot shown in **Figure 3** depicts all effect sizes plotted against their inverse standard errors: studies employing larger samples are higher on the y-axis. As can be seen, effect size seems to be inversely related to sample size for the incidental conditions. Egger's test for funnel plot asymmetry is significant across the incidental conditions, $z = 2.38$, $p = .017$, and also within the smaller subset of Restaurant studies, $z = 3.16$, $p = .002$. While this might be indicative of publication bias, other factors can cause such effects (Sterne et al., 2011) and the number of studies entering into the meta-analysis is small.

Note that in the presence of selective publication and reporting biases, meta-analytic averages can overestimate, sometimes grossly, the true mean effect (Kvarven, Strömmland, & Johannesson, 2019; Vosgerau, Simonsohn, Nelson, & Simmons, 2019). Thus the average effect of

$d = 0.19$ should probably be seen as an optimistic estimate. The suspicion that previous studies on incidental anchoring may be biased receives further support from the unusual distribution of significant p values in this set of studies. As can be seen in **Figure 3**, the effects published in previous studies all lie very close to the shaded boundary at $p = .05$. If incidental anchoring is a true phenomenon and these are adequate tests of it, then there should be more small p values than ones close to .05. To test whether the distribution of effect sizes is consistent with a true effect, we conducted a p -curve analysis (Simonsohn, Nelson, & Simmons, 2014). For this analysis we used as input all significant ($p < .05$) results in the positive direction, that is, the three anchoring effects in the original studies plus the effect for the iPad item in Study 1. For the three original studies, the p values were computed from the reported statistics in their main text. **Figure 4** shows that far from being right-skewed, the p -curve is relatively flat and, if anything, slightly left-skewed (no small p values). A binomial test shows that the relative proportion of p values above and below .025 is flatter than would be expected under the (default) null hypothesis of 33% power, although this trend is only marginally significant, $p = .078$. A continuous test against the same null hypothesis returns a just significant result, $z = -1.64$, $p = .050$. Most interestingly, p -curve analysis suggests that the average power of these studies after correcting for publication bias is just .05, as would be expected under the hypothesis that all effects are false positives. This is indicative of a set of results that lack evidential value.

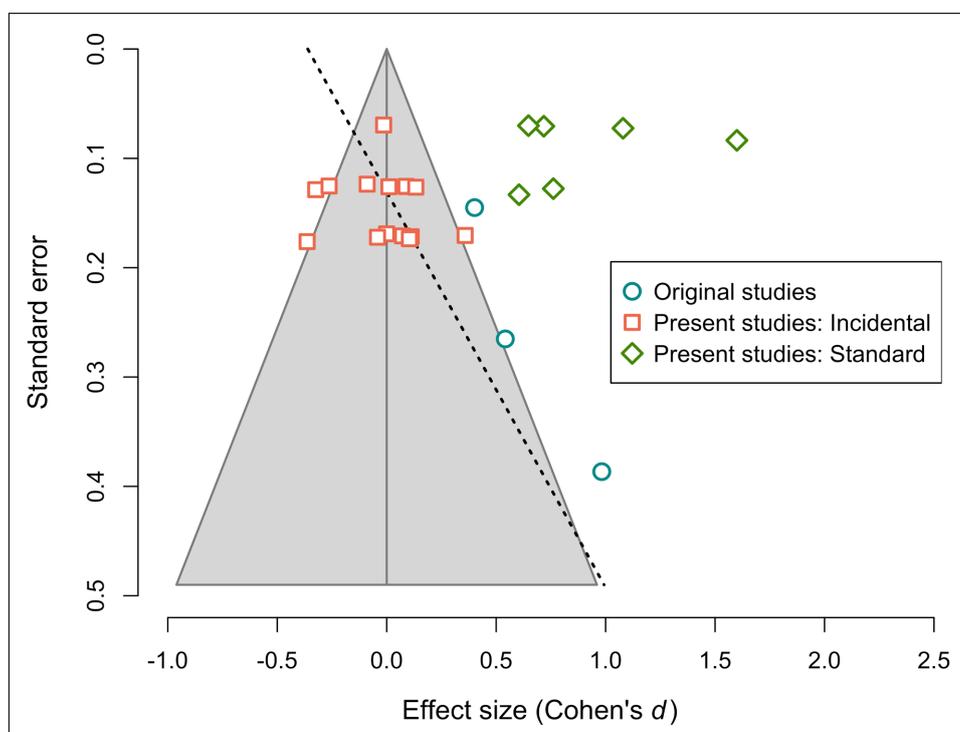


Figure 3: Funnel plot. Symbols represent the effect sizes (Cohen's d) of each of the original incidental anchoring studies (blue circles), the incidental conditions of Studies 1–3 (red squares), and the standard anchoring conditions of Studies 2 and 3 (green diamonds) plotted against the inverse of that study's SE. The dotted line is the regression line from the Egger test (only incidental anchoring studies). The shaded gray area depicts the region in which $p > .05$ for individual studies.

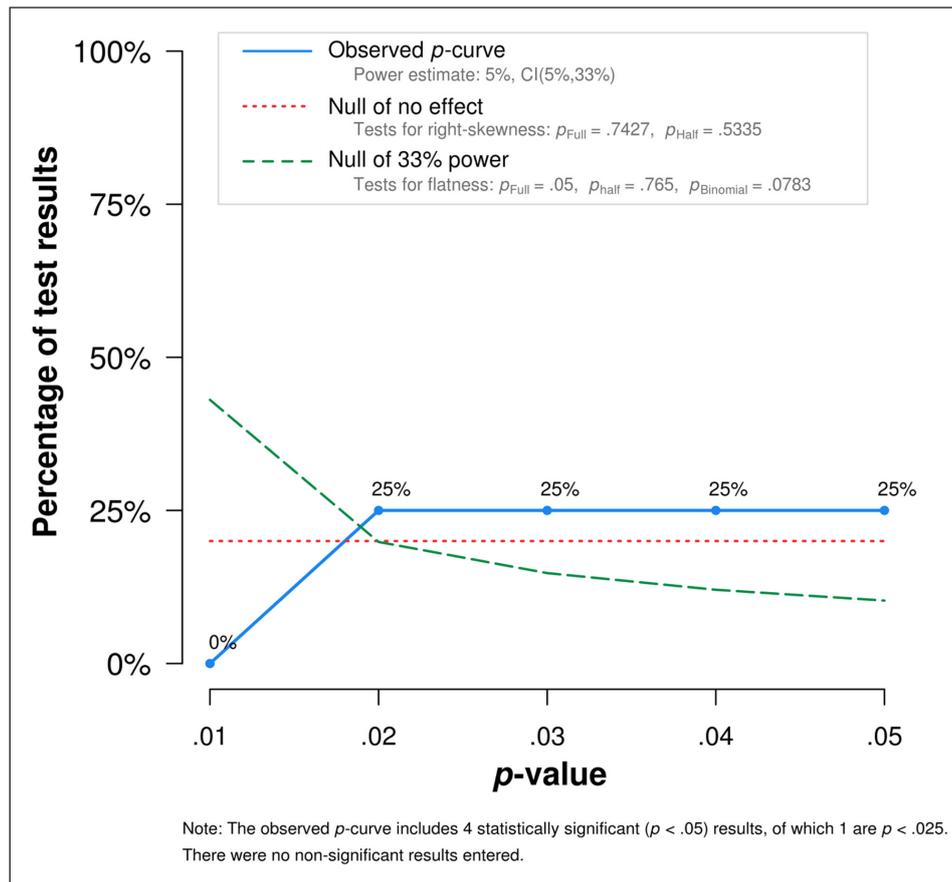


Figure 4: p-curve. Percentage of test results yielding different p values, based on all available significant ($p < .05$) results from the original studies plus the effect for the iPad item in Study 1 (solid blue line). Also depicted is the expected percentage under the null hypothesis of no effect (red dotted line) and under the null hypothesis of 33% power (green dashed line). The disclosure table for this figure is available at osf.io/8ynwu.

Overall the meta-analysis confirms that the null hypothesis (that incidental anchoring does not occur) cannot be rejected on the available evidence, but it also hints at other factors at play: These could include publication bias (null results exist but have not been published) and/or some as-yet-unknown but important moderator of incidental anchoring exists. Despite the efforts made here (especially in Study 3) to closely replicate former studies, the possibility cannot be excluded that incidental anchoring occurs under some circumstances.

General Discussion

The results reported here suggest a negligible effect of incidental environmental anchors on judgments of value. Across three high-powered studies, incidental numbers did not generate consistent effects on consumers' price estimations. These null findings, supported by Bayesian analyses of the relative support for the null hypothesis, encompass a range of goods and services including technology items, holidays, and entertainment. Of particular note is that in all three studies the first item presented to participants was the Restaurant one, hence ensuring that responses could not be contaminated or influenced by any prior estimations. Thus our procedure represents a close (near exact in Study 3) replication of the equivalent studies by Critcher and Gilovich (2008),

Dogerlioglu-Demir and Koçuş (2015), and Koçuş and Dogerlioglu-Demir (2020) but nevertheless failed to detect an incidental anchoring effect. In contrast, the results regarding the standard anchoring questions are fully in line with numerous previous demonstrations of its robustness in both the one-step and two-step procedure.

It must be acknowledged that our standard anchoring questions did not relate to price estimates and so it is not legitimate to directly compare the magnitude of the incidental and standard anchoring effects obtained. Rather, the inclusion of these items allows us to conclude that under the particular conditions of our experiments (participant samples, remuneration, etc.), we are able to obtain standard anchoring. This will reassure readers who might wonder if our null conclusions are specific to incidental anchoring or extend to all forms of anchoring. Clearly they are specific. It is noteworthy that standard anchoring on WTP estimates is well-established (Yoon et al., 2019). We also highlight the value in future research of devising versions of incidental and standard anchoring items that are much more similar – for example, using images and measuring WTP estimates in both cases – to try to more clearly pin down what the critical ingredient is that generates standard but not incidental anchoring.

The meta-analysis confirms and quantifies our overall conclusions, as well as demonstrating a small-study effect

(effect size is inversely correlated with sample size) and a lack of evidential value (p -curve) in previously-published studies. In the Introduction we noted that the statistically significant effect in at least one of the published studies (Dogerlioglu-Demir & Koçaş, 2015, Study 1) does not withstand a sensitivity analysis (Steeegen, Tuerlinckx, Gelman, & Vanpaemel, 2016) in that inclusion or exclusion of outliers alters the inference.

What theoretical rationale is there for anchoring to be strong in standard anchoring procedures and non-existent or at least very small in incidental anchoring questions? An obvious hypothesis is that the effect is a function of the level of effortful and deliberate thinking about the anchor, and in particular, that effortful thinking is a necessary condition for obtaining anchoring. In the 1- and 2-step procedures, as well as in other typical anchoring tasks, participants are required to explicitly think about or reflect on the anchor in some way. In the 2-step procedure, for example, they judge in the comparative question whether the true value is greater or less than the anchor value. In the incidental anchoring procedure, in contrast, no deliberate thinking about the anchor is required and indeed for many participants it may barely be noticed.

Over the years this idea has been rejected by many investigators (e.g., Epley & Gilovich, 2005; Kahneman, 2011, ch. 11) but we argue that this rejection is premature (see Lieder, Griffiths, Huys, & Goodman, 2018; Newell & Shanks, 2014, for fuller discussion). Evidence that anchoring is an automatic (or System 1; Kahneman, 2011) process has come, for instance, from demonstrations that it is largely immune to increased motivation to be accurate, induced by financial incentives (Chapman & Johnson, 2002), but more recent work has led to revision of this conclusion (Simmons, LeBoeuf, & Nelson, 2010). Similarly, if deliberative processes play a key role, then an expert who knows a great deal about a domain ought to be able to dilute the effect of an anchor by deliberately accessing relevant knowledge. Conversely, if anchoring is as strong in experts as in non-experts, this would imply that it is driven by automatic (System 1) processes. Just as with studies on incentives, several early reports suggested no effect of expertise (e.g., Northcraft & Neale, 1987), but more recent research challenges this conclusion (Smith, Windschitl, & Bruchmann, 2013). Lieder et al. (2018) showed that many of the benchmark properties of anchoring can be explained by a deliberative, rational resource model.

Another hypothesis – not necessarily incompatible with the one above – is that participants have to perceive that the anchor is relevant to the product attribute being evaluated (Yan & Duclos, 2013) and be aware that the anchor number is on the same judgment scale as their response estimate for anchoring to occur (but see Harris & Speekenbrink, 2016). An anchor mentioned incidentally (“Studio 17 is a new restaurant”) is unlikely to be interpreted as relevant to a price scale.

In sum, the failure to obtain incidental anchoring observed here is in line with the hypothesis that deliberate thinking about the anchor is a prerequisite for obtaining anchoring effects on judgments.

One limitation of this work is that we did not explore the range of the anchors used. For the key Restaurant item, we followed previous researchers in assessing the effects of the numbers 17 and 97. While this permitted us to investigate the replicability of the incidental anchoring effect – our major aim – it clearly limits us in terms of the more general claims we can make about the elusiveness of incidental anchoring. In studies of standard anchoring, it is well-known that the percentile gap between the anchor values (that is, where the low and high anchors fall in the distribution of judgments that participants make in the absence of any anchor) strongly moderates the magnitude of anchoring (Jung et al., 2016): Specifically, the larger the gap, the bigger the anchoring effect. What is the gap in our studies? Since Study 3 had the largest sample, we calculated the percentile rank of the low (12) and high (99) anchors against the overall distribution of WTP estimates (shown in **Figure 1**). Although the gap ($99 - 12 = 87$) would be expected to yield a very large effect in standard anchoring according to the analysis by Jung et al. (2016, see their Figure 3), future work could usefully ask whether incidental anchoring becomes detectable with different gaps.

Another major limitation of the current research is that we studied hypothetical rather than real price judgments and our findings do not rule out the possibility that incidental anchors might have an influence when people make more consequential decisions. Although some research has documented incidental anchoring effects in field studies or other applied settings, the available evidence is little more than suggestive. For example, in a well-known experiment by Nunes and Boatwright (2004, Study 1) shoppers bid more for a CD at a makeshift stand on a boardwalk when an adjacent stand was selling sweatshirts for \$80 compared to \$10, but this effect provides only ‘anecdotal’ evidence on a Bayesian analysis ($BF_{10} = 1.45$). In their third study Nunes and Boatwright found that buyers of classic cars in an auction bid more when the previous car had sold for a higher rather than lower amount. Although they included several potential moderators in their analysis, this result is correlational and hence possibly driven by some unknown confounding variable. Bobinac (2019) asked patients with chronic kidney disease to state how confident they were in their WTP estimates for a hypothetical new dialysis treatment, and concluded that these ratings correlated with the value of an irrelevant anchor number displayed at the bottom of the questionnaire. However the correlation only achieved $p = .07$. Clearly more evidence is needed about such phenomena, particularly from high-powered preregistered studies.

Conclusions

This article addresses a straightforward claim, namely that incidental ‘anchors’ such as a number in the name of a restaurant or in a product’s model name can affect how much people are willing to spend. While ‘standard’ anchoring is a very well-established phenomenon, there is much less evidence for incidental anchoring beyond the original and highly-cited article that introduced the

idea (Critcher & Gilovich, 2008) and a small number of follow-up reports. Incidental anchoring potentially has very wide relevance, even wider than standard anchoring, and could potentially influence almost any price estimate or willingness-to-pay decision.

We report 2 large initial experiments which found no evidence of incidental anchoring. The third experiment was pre-registered with a very large sample size (nearly 900 participants) but again found no hint of the effect, despite using a method very close to that of Critcher and Gilovich. By contrast our experiments find strong evidence of standard anchoring. We then describe a meta-analysis which reveals that anchoring type (incidental vs. standard) strongly moderates the outcome, and that highlights a clear small-study effect and very low evidential value amongst the previously published studies. We conclude that there is very little support for the claim that incidental anchors can bias price estimates or willingness-to-pay judgments.

Data Accessibility Statement

Data, materials, and Transparency Report (Aczel et al., 2020) are available at osf.io/8ynwu.

Additional File

The additional file for this article can be found as follows:

- **Table A1.** Questions presented in Studies 1–3. DOI: <https://doi.org/10.1525/collabra.310.s1>

Notes

¹ In their remaining experiments (Studies 2–4), Dogerlioglu-Demir and Koçaş (2015) elicited willingness-to-pay estimates for hamburger meal advertisements as a function of the product's name (e.g., 'the 1999 burger'). Because participants tended to interpret the numbers as associated with price, these are not truly incidental in the way that a number on a sportsman's jersey is incidental. The studies in Koçaş and Dogerlioglu-Demir (2020) explored ways in which ambiguous anchor values (e.g., negative values) are interpreted.

² This is based on the 5 experiments by Critcher and Gilovich (2008) and Dogerlioglu-Demir and Koçaş (2015): unweighted mean $d = 0.48$, range 0.26–0.97.

Acknowledgements

We thank Adam Harris for advice during this project, and especially Clayton Critcher and Kivilcim Dogerlioglu-Demir for sharing relevant data and materials.

Funding Information

This research was supported by grants from the United Kingdom Economic and Social Research Council to DRS and MAV and from Comunidad de Madrid (Programa de Atracción de Talento Investigador, 2016-T1/SOC-1395) to MAV.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

DS, PB-H: conceived, designed, conducted and analysed the experiments; MAV, DS: conducted meta-analyses; all authors: wrote manuscript.

References

- Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., ... Wagenmakers, E.-J.** (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*, 4–6. DOI: <https://doi.org/10.1038/s41562-019-0772-6>
- Bakker, M., van Dijk, A., & Wicherts, J. M.** (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554. DOI: <https://doi.org/10.1177/1745691612459060>
- Bobinac, A.** (2019). Mitigating hypothetical bias in willingness to pay studies: Post-estimation uncertainty and anchoring on irrelevant information. *European Journal of Health Informatics*, *20*, 75–82. DOI: <https://doi.org/10.1007/s10198-018-0983-1>
- Brewer, N. T., & Chapman, G. B.** (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making*, *15*, 65–77. DOI: <https://doi.org/10.1002/bdm.403>
- Chapman, G. B., & Johnson, E. J.** (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *The psychology of intuitive judgment: Heuristics and biases* (pp. 120–138). Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511808098.008>
- Critcher, C. R., & Gilovich, T.** (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, *21*, 241–251. DOI: <https://doi.org/10.1002/bdm.586>
- Dienes, Z.** (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290. DOI: <https://doi.org/10.1177/1745691611406920>
- Dogerlioglu-Demir, K., & Koçaş, C.** (2015). Seemingly incidental anchoring: The effect of incidental environmental anchors on consumers' willingness to pay. *Marketing Letters*, *26*, 607–618. DOI: <https://doi.org/10.1007/s11002-014-9295-0>
- Epley, N.** (2004). A tale of tuned decks? Anchoring as accessibility and anchoring as adjustment. In D. J. Koehler & N. Harvey (Eds.), *The Blackwell handbook of judgment and decision making* (pp. 240–257). Malden, MA: Blackwell. DOI: <https://doi.org/10.1002/9780470752937.ch12>
- Epley, N., & Gilovich, T.** (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18*, 199–212. DOI: <https://doi.org/10.1002/bdm.495>
- Forsell, E., Viganola, D., Pfeiffer, T., Almenberg, J., Wilson, B., Chen, Y., ... Dreber, A.** (2019). Predicting replication outcomes in the Many Labs 2 study.

- Journal of Economic Psychology*, 75. DOI: <https://doi.org/10.1016/j.joep.2018.10.009>
- Francis, G.** (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19, 151–156. DOI: <https://doi.org/10.3758/s13423-012-0227-9>
- Frederick, S. M., & Mochon, D.** (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, 141, 124–133. DOI: <https://doi.org/10.1037/a0024006>
- Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burtin, A., ... Speekenbrink, M.** (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology: General*, 148, e30–e50. DOI: <https://doi.org/10.1037/xge0000644>
- Harris, A. J. L., & Speekenbrink, M.** (2016). Semantic cross-scale numerical anchoring. *Judgment and Decision Making*, 11, 572–581.
- Jacowitz, K. E., & Kahneman, D.** (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166. DOI: <https://doi.org/10.1177/01461672952111004>
- 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, 354–368. DOI: <https://doi.org/10.1509/jmr.14.0238>
- Kahneman, D.** (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Tversky, A.** (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511803475>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahnik, S., Bernstein, M. J., ... Nosek, B. A.** (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45, 142–152. DOI: <https://doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Nosek, B. A.** (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1, 443–490. DOI: <https://doi.org/10.1177/2515245918810225>
- Koçaş, C., & Dogerlioglu-Demir, K.** (2020). The 1 in 1,000,000: Context effects of how numbers cue different kinds of incidental environmental anchoring in marketing communications. *Journal of Business Research*, 109, 536–544. DOI: <https://doi.org/10.1016/j.jbusres.2019.01.027>
- Kvarven, A., Strömland, E., & Johannesson, M.** (2019). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*. DOI: <https://doi.org/10.1038/s41562-019-0787-z>
- LeBoeuf, R. A., & Shafir, E.** (2006). The long and short of it: Physical anchoring effects. *Journal of Behavioral Decision Making*, 19, 393–406. DOI: <https://doi.org/10.1002/bdm.535>
- Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D.** (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25, 322–349. DOI: <https://doi.org/10.3758/s13423-017-1286-8>
- Matthews, W. J.** (2011). What might judgment and decision making research be like if we took a Bayesian approach to hypothesis testing? *Judgment and Decision Making*, 6, 843–856.
- Mussweiler, T., & Strack, F.** (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136–164. DOI: <https://doi.org/10.1006/jesp.1998.1364>
- Newell, B. R., Lagnado, D. A., & Shanks, D. R.** (2015). *Straight choices: The psychology of decision making* (2nd ed.). Hove, UK: Psychology Press. DOI: <https://doi.org/10.4324/9781315727080>
- Newell, B. R., & Shanks, D. R.** (2014). Prime numbers: Anchoring and its implications for theories of behavior priming. *Social Cognition*, 32, 88–108. DOI: <https://doi.org/10.1521/soco.2014.32.sup.88>
- Northcraft, G. B., & Neale, M. A.** (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, 39, 84–97. DOI: [https://doi.org/10.1016/0749-5978\(87\)90046-X](https://doi.org/10.1016/0749-5978(87)90046-X)
- Nunes, J. C., & Boatwright, P.** (2004). Incidental prices and their effect on willingness to pay. *Journal of Marketing Research*, 41, 457–466. DOI: <https://doi.org/10.1509/jmkr.41.4.457.47014>
- R Core Team.** (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D.** (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology*, 99, 917–932. DOI: <https://doi.org/10.1037/a0021540>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P.** (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*, 143, 534–547. DOI: <https://doi.org/10.1037/a0033242>
- Simonson, I., & Drolet, A.** (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, 31, 681–690. DOI: <https://doi.org/10.1086/425103>
- Smith, A. R., Windschitl, P. D., & Bruchmann, K.** (2013). Knowledge matters: Anchoring effects are moderated by knowledge level. *European Journal of Social Psychology*, 43, 97–108. DOI: <https://doi.org/10.1002/ejsp.1921>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W.** (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11, 702–712. DOI: <https://doi.org/10.1177/1745691616658637>

- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., ... Higgins, J. P. T.** (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, *343*. DOI: <https://doi.org/10.1136/bmj.d4002>
- Strack, F., & Mussweiler, T.** (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*, 437–446. DOI: <https://doi.org/10.1037/0022-3514.73.3.437>
- Tversky, A., & Kahneman, D.** (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. DOI: <https://doi.org/10.1126/science.185.4157.1124>
- Viechtbauer, W.** (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*. DOI: <https://doi.org/10.18637/jss.v036.i03>
- Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P.** (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General*, *148*, 1628–1639. DOI: <https://doi.org/10.1037/xge0000663>
- Werthenbroch, K., & Skiera, B.** (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research*, *39*, 228–241. DOI: <https://doi.org/10.1509/jmkr.39.2.228.19086>
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N.** (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387–402. DOI: <https://doi.org/10.1037/0096-3445.125.4.387>
- Yan, D., & Duclos, R.** (2013). Making sense of numbers: Effects of alphanumeric brands on consumer inference. *International Journal of Research in Marketing*, *30*, 179–184. DOI: <https://doi.org/10.1016/j.ijresmar.2012.09.007>
- Yoon, S., Fong, N. M., & Dimoka, A.** (2019). The robustness of anchoring effects on preferential judgments. *Judgment and Decision Making*, *14*, 470–487.

Peer review comments

The author(s) of this paper chose the Open Review option, and the peer review comments can be downloaded at: <http://doi.org/10.1525/collabra.310.pr>

How to cite this article: Shanks, D. R., Barbieri-Hermitte, P., & Vadillo, M. A. (2020). Do Incidental Environmental Anchors Bias Consumers' Price Estimations? *Collabra: Psychology*, *6*(1): 19. DOI: <https://doi.org/10.1525/collabra.310>

Senior Editor: Don Moore

Editor: Don Moore

Submitted: 24 December 2019

Accepted: 03 March 2020

Published: 03 April 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.