VRIJE
UNIVERSITEIT
BRUSSEL

Master thesis submitted in partial fulfillment of the requirements for the diploma Master of Science in TEW: Handelsingenieur

GOTCHA WITH A CAPTCHA! Basic anchoring effects and willingness to pay in an e-commerce setting

Sebastiaan MICHIELS

0505322

Academic year 2017-2018

Promotor: Leo VAN HOVE

Economic & Social Sciences & Solvay Business School

Basic anchoring effects and willingness to pay in an e-commerce setting

Sebastiaan Michiels
0505322

VRIJE UNIVERSITEIT BRUSSEL

Promotor: Professor Leo Van Hove

Vrije Universiteit Brussel, May 22nd, 2018

Abstract

The influence of anchoring effects on price judgments has primarily been studied for relevant anchors, called reference prices, but far less for irrelevant anchors. Although standard anchoring effects have shown their robustness in many studies, the inherent comparative judgment limits their usefulness in business applications. Past demonstrations of a basic anchoring effect, however, uncovered the potential impact of irrelevant anchors in the absence of a comparative judgment. Based on the anchoring literature and Thaler's (1985) positive theory of consumer choice, I propose a practical application that uses basic anchoring effects to influence the willingness to pay of consumers in an e-commerce setting. An experiment (N=232) demonstrates that numeric CAPTCHAs can influence price judgments provided that there is sufficient mental processing of the anchor value. The resulting basic anchoring effects turn out to be at least as sizable as standard anchoring effects. These findings suggest that basic anchoring effects have a broader applicability in business settings than standard anchoring effects.

Keywords: basic anchoring effect; consumer choice; e-commerce
JEL classification: D12, D91

Table of Contents

Author's note

In 2006, when I was eleven years old, I got the opportunity to write a children's book together with a successful author. Three years later a few hundred copies of the book were sold. Time passed without thinking about this experience until I read a paper by Mussweiler and Englich (2005) in which they use subliminal visual cues to influence judgments. In the book I co-wrote, a smart but malicious owner of a retail store uses psychological tricks to influence consumers to buy what they do not need. One of these tricks is the use of subliminal visual cues by inserting frames in TV advertisements. Since the frames are only shown for a couple of milliseconds the viewer has not consciously seen them, even though they are processed by his brain. I believed it to be a coincidence that I ended up writing my master dissertation about another psychological trick, the anchoring effect. Thinking back at the book and its store owner made me reconsider. If I had had knowledge of the power of anchoring effects, it too would have been part of the store owner's weaponry. However, with this master dissertation I do not aim to arm firm owners more than I want to inform and warn consumers.

This text is not written as a journal article, in order to be understandable for a lay person in the field of anchoring effects and price judgments. Almost every academic paper cited in this dissertation uses only a couple of sentences to explain the anchoring effect. Most of the conversations I have had with students or researchers taught me that such a concise explanation is not enough to make lay people understand what anchoring effects are, and what they are not. Therefore, writing this dissertation was a juggling act in which I attempted to reconcile completeness, conciseness, and understandability. Some of the paragraphs are based (sometimes verbatim) on a paper I wrote in 2016-2017 to elucidate the subject of my master dissertation[1].

Although anchoring effects are mainly studied from a (cognitive) psychological point of view, I lean more upon (behavioral) economics than the majority of the researchers. As a prospective Master of Science in Business Engineering, I deemed discussing the possibility of a business application involving basic anchoring effects appropriate.

[1] The most important paragraphs can be found in sections 2.1-2.3, 2.5, 3.1, 3.2.1, 4.1, 5.1, and 5.4.

When I proposed this research subject to Professor Leo Van Hove, who eventually became the promotor for my master dissertation, he responded straightaway with enthusiastic commitment and useful comments. I would like to express my special thanks to Professor Leo Van Hove for his fascinating course on IT Economics, for the witty title he came up with, and for the other valuable suggestions that have without doubt improved my master dissertation. I also wish to acknowledge the help provided by his assistant, drs. Ellen Van Droogenbroeck.

My appreciation also goes to Professor Bruno Heyndels, who aroused my interest in behavioral economics during his Microeconomics course. Thanks to him I became familiar with Amos Tversky and Daniel Kahneman, which in turn made me pick up Kahneman's book 'Thinking, Fast and Slow' when I noticed it at the library by chance. After this turning point, I read a few more books on behavioral economics and cognitive psychology. 'Thinking, Fast and Slow', 'Misbehaving', and 'Predictably Irrational' have directly and indirectly influenced this master dissertation and have proven to be useful in everyday life. For this I am extremely grateful to their respective authors Daniel Kahneman, Richard Thaler, and Dan Ariely.

Writing a master dissertation is not only a juggling act, but also a marathon. As the proverb says: if you want to go far, go together. I dedicate this dissertation to my parents and my closest friends, for their unconditional love and support.

Sebastiaan Michiels
May 22nd, 2018

1 Introduction

Consider the following anecdotes:

1. Mr. A and Mr. B are seasoned real estate agents, who separately visit a property. They have received an extensive document with information about the property and other properties in the same neighborhood. Due to a mistake, the booklet of Mr. A contains a listing price of $65,900, which is different from the $83,900 in Mr. B's booklet. After the visit, Mr. A and Mr. B estimate the appraised value of the property at $68,000 and $75,000, respectively. They insist that their estimations are not based on the provided listing price.

2. Two lawyers, Ms. K and Ms. L, participate in an experiment in which they both act as a judge in a fictitious shoplifting case. Before determining the final sentence, they are asked to examine realistic case material and evaluate whether the sentencing demand of the prosecutor is too low, too high, or just right. Because it is an experiment, Ms. K and Ms. L determine this sentencing demand randomly by throwing a pair of dice and calculating the sum of the two dice. This leads to a prosecutor's demand of 3 months for Ms. K and 9 months for Ms. L. The shoplifter is sentenced to 5 months on probation by Ms. K. Surprisingly, Ms. L imposes a sentence of 8 months on probation, even though they both evaluated the same crime.

3. Mr. and Mrs. X want to buy a car. Before they are asked to estimate the average price of a midsize car, they focus sixty seconds on a computer screen that shows a meaningless letter string. Mr. and Mrs. X look at different monitors, but the letter string is the same. It is not by coincidence, however, that Mr. X provides a lower estimate (€17,000) than his wife (€21,000).

Up to this point, one could attribute this peculiar anecdotal evidence to coincidence or to differences between the pairs with regard to their knowledge, preferences, et cetera. However, the anecdotes are based on the results of renowned empirical studies[1]. Every character is the personification of the average decision maker of a bigger group[2]. Ms. K, for instance, represents the group of lawyers who threw '3' and imposed an average sentence of 5.28 months.

[1] The anecdotes are based on papers by Northcraft and Neale (1987; Experiment 1), Englich, Mussweiler, and Strack (2006; Study 3), and Mussweiler and Englich (2005; Study 2), respectively.
[2] In fact, Mr. A and Mr. B represent the *median* decision maker since their final estimates are the median estimates in the experiment by Northcraft and Neale (1987).

It is hardly a surprise that amateurs are influenced by the listing price of a house. However, in a paper on "Experts, Amateurs, and Real Estate", Northcraft and Neale (1987) demonstrate that experienced real estate agents are almost as susceptible to the listing price as amateurs. It turns out that experts and amateurs differ in only one aspect. While amateurs acknowledged the important role of the listing price in their judgment process, experts denied any influence. Is it possible that the listing price affects Mr. A's and Mr. B's judgment, even though they are not aware of its influence?[3]

"Playing dice with criminal sentences" is the title of a famous paper by Englich, Mussweiler, and Strack (2006), and an apt description of Ms. K's and Ms. L's behavior. Every lawyer is well aware that the sum of two dice is unrelated to the appropriate sentence for a shoplifter, even more so because the instructions of the experiment explicitly mentioned that the randomly determined prosecutor's sentence demand "does not reflect any juridical expertise" (Englich et al., 2006, p. 192). How could a single throw of dice make a difference of three months for the shoplifter? One could argue that Ms. K and Ms. L can only be 'tricked' by a number (3 or 9) – which they obviously should have ignored – because they are participating in an experiment. If their sentences had had real consequences for the shoplifter, the lawyers would not have been susceptible to the influence of the dice. Unfortunately, a study by Ariely, Loewenstein, and Prelec (2003) provides strong empirical evidence against the hypothesis that raising the stakes eliminates anomalies.

During a lecture, Ariely et al. (2003) put several identical bottles of wine up for auction. First, they asked the students to write down the last two digits of their social security number (SSN) and to indicate whether they would buy a bottle of wine for a price equal to the number formed by these digits. Later, the students stated the maximum price they would be willing to pay for the product. This experiment had real consequences – some students ended up having to buy a bottle of wine – but the SSNs nevertheless influenced the bids for identical bottles of wine. Student with a low-ending SSN (from 01 to 20) bid on average €12. Students with a high-ending SSN (from 80 to 99) were willing to pay €38 for the same bottle of wine. How can an SSN influence bids in a real auction?

[3] The present paper will gradually answer all introductory questions.

On the face of it, the third anecdote – which I based on a study by Mussweiler and Englich (2005) – seems hardly related to the other two. In contrast to the lawyers and the students, Mr. and Mrs. X are not asked to compare their best guess for the average price of a midsize car with another numeric value. In fact, Mr. and Mrs. X do not even perceive a numeric value. There is no listing price, they do not throw a pair of dice, and they certainly do not write down the last two digits of their SSN. However, Mr. and Mrs. X were exposed to subliminal visual cues while focusing on the meaningless letter string. A frame with a number (10,000 for Mr. X and 30,000 for his wife) was shown multiple times, but only for a couple of milliseconds. Mr. and Mrs. X have not *consciously seen* the number, but it has nonetheless been processed by their brains, and led to different estimates. When decision makers like Mr. and Mrs. X do not realize that a number influences their judgments, should we too become very suspicious of the large amount of numbers we encounter every day?

The anecdotes illustrate how numeric values, which are called 'anchors', can affect judgments in a predictable way: a high (low) anchor leads to a higher (lower) estimate. All characters, from Mr. A to Mrs. X, have become subject to so-called 'anchoring effects'.

Wong and Kwong (2000) highlight three main reasons why studying the anchoring phenomenon remains valuable, even after several decades of academic research. First of all, the anchoring effect potentially has many applications in the real world. Secondly, as has been illustrated by the anecdotes, it can affect all our day-to-day judgments, from answering factual questions to spousal preference (Davis, Hoch, & Ragsdale, 1986). Finally, a better understanding of the anchoring effect contributes to the research on related psychological phenomena such as hindsight bias (see Fischhoff, 1975). I add to these reasons the need for a decent literature review, because since the turn of the century the research domain has become increasingly diverse and fragmented. The anecdotes indicate that there are several types of anchoring effects (e.g. based on the type of anchor, the way the anchor is presented, etc.) and some of the more recent demonstrations question the established theories and frameworks.

The contribution of the present dissertation to the existing literature is multidimensional. In search of a practical application that uses 'basic anchoring effects'[4] to influence the willingness to pay of consumers in an e-commerce setting, existing theories and frameworks are tested to widen the knowledge of the anchoring effect. At the same time, I propose some new frameworks to bring together the scattered anchoring literature.

Chapter 2 provides an overview of the anchoring literature and builds an understanding of anchoring effects in general, and basic anchoring effects in particular. Chapter 3 discusses the literature related to price judgments and consumer choice. After the critical literature review in chapters 2 and 3, chapter 4 sets out a practical application to influence the willingness to pay of consumers, based on basic anchoring effects. The potential for this application is tested in an experiment. The experiment and its main results are presented in chapters 5 and 6. Chapter 7 highlights the theoretical and managerial implications. Chapter 8 contains the final conclusion, which I already gave away in the title: 'Gotcha with a CAPTCHA[5]!'.

[4] Basic anchoring effects are a subset of anchoring effects (see section 2.4 for a classification of the different anchoring effects). The effects described in the first and second anecdote are not basic anchoring effects. The effect demonstrated in the third anecdote is a (special type of) basic anchoring effect.
[5] Gotcha is colloquial English and means here 'I got you by surprise'. A CAPTCHA is a tool used by websites to verify that they are dealing with human users (see section 4.3).

2 Anchoring effects

This chapter provides a critical overview of the anchoring literature. First, the origin and nature of the anchoring effect are discussed in sections 2.1 and 2.2. The introductory anecdotes illustrated that some anchors are more relevant than others. The influence of a listing price, for instance, is less surprising than the impact of an SSN. Therefore, the different anchor types are categorized in section 2.3. Section 2.4 examines how anchoring effects are measured and how an experimenter can determine optimal anchor values. At the end of the chapter, in section 2.5, we are able to define the concept of basic anchoring effects.

2.1 Heuristics and biases

Tversky and Kahneman[6] (1974) commonly receive credit for the discovery of the anchoring phenomenon, which they named the "adjustment and anchoring" heuristic (p. 1228). Their paper (Tversky & Kahneman, 1974) was not only the starting point for the literature on anchoring, but also the genesis of a whole research domain on heuristics and biases in judgment and decision making (Costa, de Melo Carvalho, de Melo Moreira, & do Prado, 2017). Before examining the adjustment and anchoring heuristic in section 2.2, a basic understanding of heuristics and bias is required.

Judgmental heuristics are simple cognitive procedures that "reduce complex tasks . . . to simpler judgmental operations" (Tversky & Kahneman, 1974, p. 1124). One could think of them as mental shortcuts. They reduce effortful operations and are in some cases even the only way to obtain a satisfactory outcome. Although heuristics usually increase both the efficiency and effectiveness of judgmental processes, blindly relying on heuristics sometimes leads to "severe and systematic errors" (Tversky & Kahneman, 1974, p. 1124)[7].

This particular finding has played a pivotal role in the making of behavioral economics. One of the reasons why standard economic theory focuses on rational agents, rather than human beings and their flaws, is the idea that human errors are distributed just so that they cancel each other

[6] Daniel Kahneman received the Nobel Memorial Prize in Economic Sciences (short for 'The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel') in 2002 for his contributions to the integration of psychology into economics. His longtime collaborator, Amos Tversky, has been as important for the field of (behavioral) economics, but he passed away in 1996.

[7] Nevertheless, heuristics are here to stay, since "in the long run . . . betting those shortcut odds may represent the most rational approach possible" (Cialdini, 2007, p. 6).

out. However, the argument does not hold water because the errors are similar[8]. This results in systematic – and to a large extent predictable – deviations from optimal judgments. These deviations are called biases.

Because of the continuous use of heuristics in a wide range of situations and the predictability of the resultant errors, cognitive biases predominate the research on judgment and decision making. Studying these biases is interesting for decision makers who want to 'debias' their own judgments, but also for everyone who wants to exploit the flawed decision making of others. At the end of this paper, in chapter 7, I discuss how anchoring effects can be used to help decision makers, but also how sellers can use it to take advantage of their consumers' errors[9]. Unfortunately, debiasing is particularly difficult in the case of anchoring effects (see section 2.5.4). But before this can be discussed, we have to answer an essential question: what are anchoring effects?

2.2 Anchoring

The literature on anchoring has evolved dramatically, but different views on anchoring still exist, including the oldest one: the adjustment and anchoring heuristic. Therefore, a section that examines the adjustment and anchoring heuristic precedes section 2.2.2, which presents a much broader view on anchoring.

2.2.1 Adjustment and anchoring heuristic

Slovic and Lichtenstein (1971) were the first to identify the impact of (insufficient) adjustment from (arbitrarily chosen) starting points on subsequent estimates. Although this represents the basic idea behind the adjustment and anchoring heuristic, it was an experiment by Tversky and Kahneman (1974, p. 1128) that made anchoring famous.

[8] Consider for instance the example of Northcraft and Neale (1987). If the errors had canceled each other out, there would have been as many decisions makers who underestimated the appraised value of a house – in spite of a high listing price – as there are decisions makers who overestimated this appraised value due to the high listing price.
[9] The Save More Tomorrow™ program (Thaler & Benartzi, 2004) is a famous example of how the understanding of the 'present bias' (also called 'hyperbolic discounting') allows the design of a choice architecture that is beneficial to the decision maker. "Goldilocks pricing" – adding a decoy premium version to a product line – is an example of how sellers can take advantage of their consumers (Varian & Shapiro, 1999, p. 70; based on Simonson & Tversky, 1992), in this case by exploiting the asymmetric dominance phenomenon described by Huber, Payne, and Puto (1982).

At the beginning of the experiment, Tversky and Kahneman spun a wheel of fortune. The wheel was marked from 0 to 100, but it was rigged to stop only at 10 and 65. Tversky and Kahneman demanded the participants to write down the obtained number and asked them whether they estimated the percentage of African countries in the United Nations to be larger or smaller than this number (10 or 65). Subsequently, the participants had to provide an estimate for the percentage of African countries in the United Nations. Although it is absurd that participants assign credibility to a number obtained by spinning a wheel of fortune, its effect was substantial. The median estimates of the participants who received 10 and 65 were 25% and 45%, respectively (Tversky & Kahneman, 1974).

The name 'adjustment and anchoring heuristic' is illustrative for the mechanism that Tversky and Kahneman (1974) considered as a possible psychological account for the phenomenon. Since the 'target value', the percentage of African countries in the United Nations, is unknown to most participants, they resort to a heuristic procedure. Estimates often start from a reference point, i.e. the anchor values (10 and 65), and are adjusted in the direction of the target value (Tversky & Kahneman, 1974). The adjustment from anchors is typically insufficient (Slovic & Lichtenstein, 1971). This made Tversky and Kahneman (1974) conclude that different anchors yield different estimates, a textbook case of a heuristic leading to a biased outcome.

Figure 1 illustrates how the adjustment and anchoring heuristic leads to an anchoring effect for participants in the high and low anchor condition, respectively. Figure 2 combines both conditions to provide an overview of the mechanism behind the adjustment and anchoring heuristic. Although section 2.4 will discuss how anchoring effects are measured, I already want to highlight the difference between the anchoring effect (i.e. the deviation of the final estimate from an unanchored estimate[10]) and the bias (i.e. the deviation of the final estimate from the target value). Figure 3 illustrates this difference and provides an indication that anchoring effects can also be used to aid decision makers with a plausible range that does not even contain the target value. This difference is, however, not made in the literature on the adjustment and anchoring heuristic. The assumption is implicitly made that the unanchored estimate equals the target value.

[10] Figure 2 suggests that the unanchored estimate lies in the middle of the plausible range, but this need not be the case. The figures are only illustrations. The insufficient adjustment, for example, can also stop at the boundaries of the plausible range.

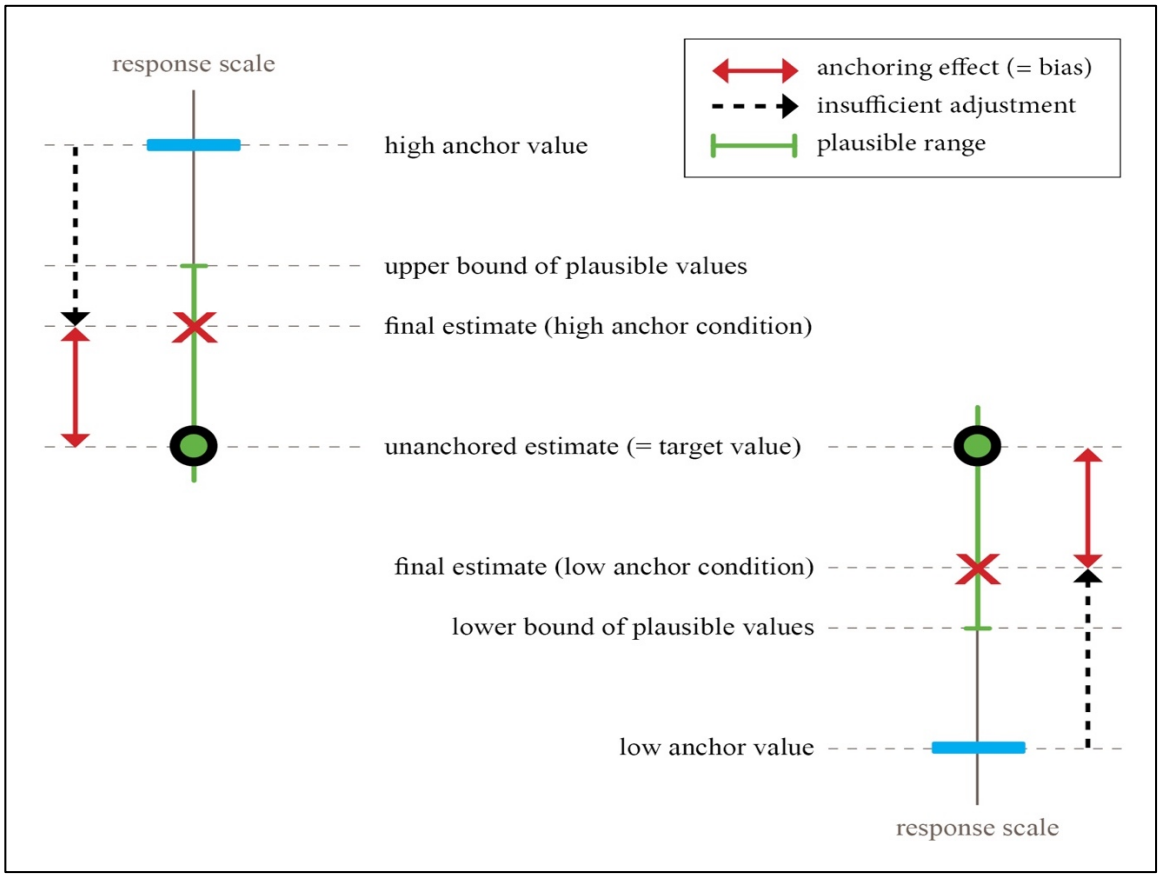


Figure 1. Illustration of adjustment and anchoring in the high and low anchor condition.

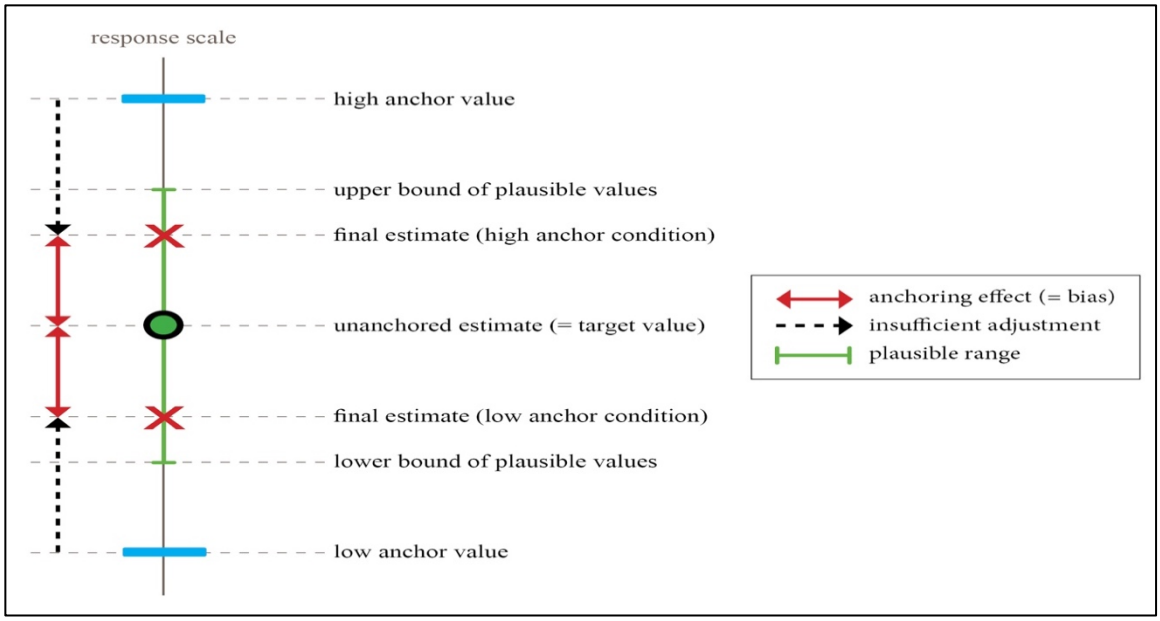


Figure 2. Illustration of the adjustment and anchoring account. Adapted from "Exploring anchoring effect and the moderating role of repeated anchor in electronic commerce," by C. S. Wu, F. F. Cheng, and H. H. Lin, 2008, *Behaviour & Information Technology*, *27*(1), p. 33. Copyright 2008 by Taylor & Francis.

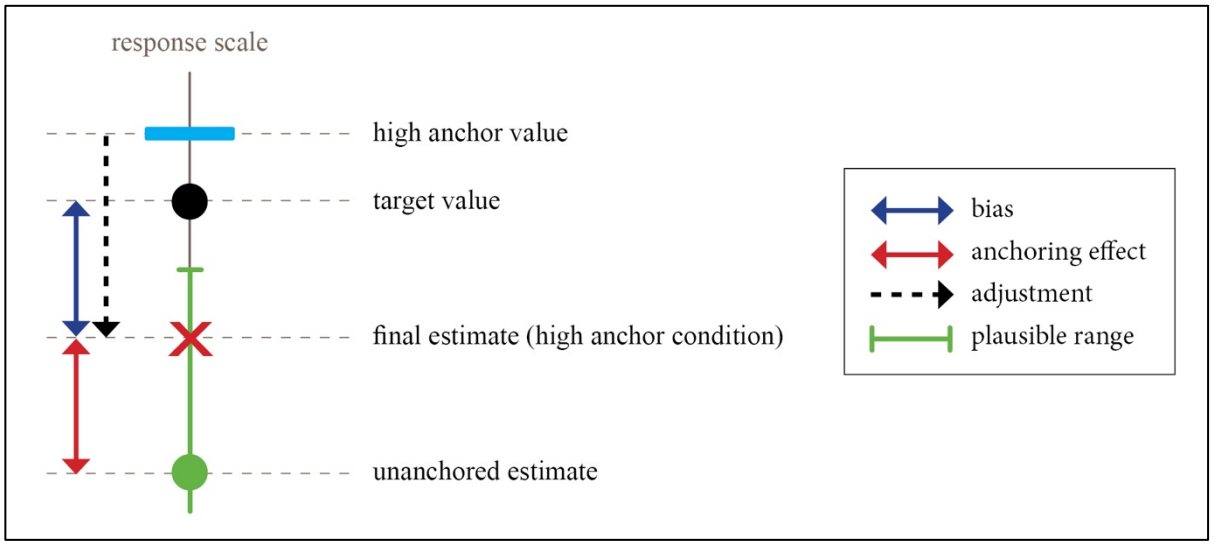


Figure 3. Bias versus anchoring effect in the adjustment and anchoring heuristic.

2.2.2 Anchoring effect

During the last decades an impressive number of studies created a fragmented body of literature on a much more complex phenomenon than described by Tversky and Kahneman (1974). Anchoring and adjustment, which is deemed to be a rather deliberate and effortful process, turned out to be only one of the possible accounts of what Wilson, Houston, Etling, and Brekke (1996) started to call an assimilation effect due to anchoring. The adjustment account cannot describe, for instance, how anchor values within the boundaries of the plausible range affect judgments (Strack & Mussweiler, 1997). Neither can it explain the impact of subliminal anchoring (Mussweiler & Englich, 2005; see chapter 1). Most demonstrations of basic anchoring effects (see section 2.5.1) are explained by numeric priming (Mussweiler & Englich, 2005; Wilson et al. ,1996; Wong & Kwong, 2000) [11]. Even though not all of the currently known underlying mechanisms were discovered at the time, Strack & Mussweiler (1997) already argued that the occurrence of different psychological mechanisms made anchoring rather an effect than a heuristic. Anchoring effect became the prevailing name[12].

Quite remarkably, many definitions of the anchoring effect (i.a. Jacowitz & Kahneman, 1995; Kahneman, 2011; Wu, Cheng, & Lin, 2008) still include references to insufficient adjustment and/or the traditional experimental paradigm, which forces the decision maker to consider the

[11] Numeric priming is a rather broad account since it states that any number – "regardless of its source, context, scale, or relevance" – can influence judgments (Mochon & Frederick, 2013, p. 70). Therefore, it is only one of the coexisting underlying mechanisms, of which the most important are anchoring and adjustment, selective accessibility, numeric priming, and scale distortion). For a recent overview see Mochon and Frederick (2013).
[12] Not only for the 'effect' itself, but also for the whole phenomenon.

anchor value as a possible answer to the target value (see section 2.5 for a classification of the experimental paradigms). New underlying mechanisms and new experimental paradigms demand a general and inherently vague definition, which I base on a more recent definition by Mochon and Frederick (2013, p. 69).

The anchoring effect is the effect of a numeric value (anchor value) on the simultaneous or subsequent estimation of an unknown numeric value (target value), provided that this effect represents an assimilation of the estimate towards the anchor[13]. In short, the anchoring effect refers to the influence of a numeric stimulus on a numeric judgment, where high (low) numeric values imply higher (lower) estimates.

The definition of anchoring effects contains several degrees of freedom. It does not specify the characteristics of the anchor (i.e. relevance, numeric value, and presentation) nor the experimental paradigm. In the remainder of this chapter, these elements will be discussed.

2.3 Anchors

Anchors can hold information that is (deemed) useful to estimate the target value. If this is the case, they are labeled relevant, with relevant meaning no more or less than 'informative with regard to the target value'. Consider, for example, the listing price of a house, financial benchmarking, and suggested retail prices. Conversely, numbers obtained by spinning a wheel of fortune or throwing a pair of dice, can hardly be considered useful.

The existing literature labels anchors as (ir)relevant, (un)informative, (im)plausible, etc. but does not provide clear definitions and fails in stating the differences and/or similarities between these concepts. I propose a more thorough definition of relevant anchors based on two dimensions: context alignment and value alignment[14].

[13] This assimilation is a less effortful and deliberate process than the adjustment in the adjustment and anchoring heuristic. In theory, an anchor value can also cause a contrast effect, but the general consensus is that the term anchoring effects is only used in case of an assimilation towards the anchor.
[14] The frameworks proposed in this dissertation have three purposes: to structure the discussion, to encourage thinking about the (proposed) classifications, and to generate questions for future research.

2.3.1 Context and value alignment

Context alignment requires that the anchor is related to context of the target question[15]. This also implies that a correspondence between the dimension of the anchor and the target is necessary. The anchor value and the target value should for instance both be expressed as prices or as percentages. Value alignment, for its part, refers to the plausibility of having the numerical value of the anchor as an answer to the target question. Table 1 illustrates four possibilities for the target question – inspired by Northcraft and Neale (1987) – 'How much would you pay for this modest house?'. Note that plausibility of the anchor is tested with respect to the target value. A reasonable (plausible) price for the castle is still implausible when considering how much to pay for a modest house.

Table 1

Illustration of the framework on relevance of anchors

		context alignment Is the context of the anchor related to the target question?	
		yes	no
value alignment Is the anchor value a plausible target value?	yes	*ask price for the house*	*number retrieved from a wheel of fortune with values that would make plausible Euro values for the house*
	no	*price for an historic castle*	*number of inhabitants of France*

Note. Target question: 'How much would you pay for this modest house?'.

Based on context and value alignment, a distinction can be made between four types of anchors (see Table 2): relevant anchors, anchors that are irrelevant due to implausibility, irrelevant despite plausibility, or undoubtedly irrelevant. Unless the specific type of irrelevance adds to the discussion, in the remainder of the text the term 'irrelevant' will cover all three types.

[15] A term sometimes used to refer to the idea of context alignment is conceptual relevance.

Table 2

A framework on relevance of anchors

		context alignment Is the context of the anchor related to the target question?	
		yes	no
value alignment Is the anchor value a plausible target value?	yes	relevant	irrelevant despite plausibility
	no	irrelevant due to implausibility	undoubtedly irrelevant

When participants assess the value alignment and, to a lesser extent, the context alignment of a given anchor, this is an individual and subjective assessment. The same anchor can be categorized as relevant and irrelevant by different participants. Evaluating plausibility depends, more than the context alignment, upon the knowledge and reasoning skills of the respondent. However, the absence of a clear cut-off value that is considered plausible (or implausible) by all participants, poses no problem if researchers choose anchors that belong indisputably to a specific quadrant of Table 2. The anchors in the wheel of fortune example, for example, are expected to be 'irrelevant despite plausibility' for all participants.

2.3.2 Relevance

Initially, most research examined the effects of relevant anchors. Tversky and Kahneman (1974) were the first to demonstrate the power of irrelevant anchors (Kahneman, 2011). Later studies confirmed the occurrence of anchoring effects when using irrelevant anchors (e.g. Critcher & Gilovich, 2008; Englich & Mussweiler, 2001; Englich et al., 2006). Most of the time, the irrelevance stems from a lack of context alignment. Examples are the outcome of a wheel of fortune (Tversky & Kahneman, 1974), someone's social security number (Ariely et al., 2003), an experimenter-provided ID number (Switzer & Sniezek, 1991; Wilson et al., 1996), a randomly chosen playing card (Cervone & Peak, 1986), or the outcome of throwing a dice (Englich et al., 2006).

Most of the existing literature states that irrelevant anchors create anchoring effects of magnitudes similar to those produced by relevant anchors (Furnham & Boo, 2011), but some studies suggest that relevant anchors produce greater effects (e.g. Glöckner & Englich, 2015).

Whatever the case, there is no indication that relevant anchors lead to smaller effects. Still, relevant anchors are sometimes mistakenly perceived as less harmful than irrelevant anchors. One could argue that a decision maker will, more often than not, be aware of the influence of relevant anchors such as suggested retail prices, and therefore able to debias his judgment. However, section 2.5.4 will explain that awareness is no sufficient condition for debiasing anchoring effects (Wilson & Brekke, 1994). Even if debiasing were easy, there is no undisputed support for the increased awareness due to relevant anchors. Northcraft and Neale (1987), for instance, show how real-estate experts were significantly influenced by the ask price of the house, but did not mention the ask price when asked which factors had influenced their judgment. Last but not least, relevance does not imply that the anchor is helpful or trustworthy. An ask price of €500,000 could be considered plausible by an individual even though the house is worth far less. Insufficient downward adjustment can cause the individual to bid too much only because of the high ask price acting as a relevant but harmful anchor.

The impact of value alignment or plausibility on the size of anchoring effects has also been examined in situations without context alignment, i.e. for anchors that are irrelevant despite plausibility and anchors that are undoubtedly irrelevant. In the literature there is a fragile consensus that using plausible anchors maximizes the chances of obtaining sizable anchoring effects. Studies show that choosing ever higher (or lower) anchor values does not produce bigger anchoring effects once the anchor values fall outside the range of plausible values (Mussweiler & Strack, 2001). Wegener, Petty, Detweiler-Bedell, and Jarvis (2001) even demonstrate curvilinear effects, implying that implausible anchors produce smaller, not bigger, anchoring effects than plausible anchors.

2.4 Measuring anchoring effects

This section discusses how the size of anchoring effects is measured and how the numeric values of plausible and implausible anchors can be determined.

2.4.1 Independent measures

It is difficult to observe anchoring effects at the level of an individual participant. Imagine asking a participant to express his best guess for the selling price of a particular house. If the

anchor[16] is provided prior to the estimation, there is no way to be sure that an anchoring effect occurred, let alone gauge its direction and size, simply because the 'anchored' estimate cannot be used to figure out what the participant's estimate would have been in a situation without the anchor. Of course, one could ask the participant to estimate twice: one 'unanchored' estimate, followed by the anchor and an anchored estimate. This procedure allows a comparison between the *unanchored* first estimate and the *anchored* second estimate. The problem is that by following this procedure an extra anchor is added to the mix. It is likely that the participant's second and anchored estimate will be influenced by his first and unanchored estimate[17]. Due to this carry-over effect there is no way to determine what the effect of the anchor would have been if one had not asked for an unanchored estimate.

For these reasons, researchers examine groups of participants rather than individuals. There are two main ways to demonstrate the occurrence of anchoring effects, and both are based on a between-subjects or independent-measures design. The first approach compares the estimates of a group of participants exposed to a high anchor with those of another group of participants exposed to a low anchor (see Figure 2). Significant differences between the estimates of the two groups would reveal the effect of the anchors. In their wheel of fortune example, Tversky and Kahneman (1974) used a high and low anchor (65 and 10, respectively). This led to a difference in the median estimates between the two groups (45 versus 25). The approach with a high and a low anchor dominated research in the early days. Since then, experiments with only one anchor started to occur more frequently (e.g. Jacowitz & Kahneman, 1995). In such a set-up, the results of a group of participants that have answered the target question without an anchor is required. The estimates of this unanchored group are then compared with a group that did receive an anchor (see Figure 1).

If the number of participants in each group is sufficiently large – with the threshold depending on the homogeneity of the population under study – and if the participants are randomly assigned to a group, one would – in the absence of anchoring effects – expect the distributions of the price estimates to be highly similar across the two groups. If, however, the dissimilarities between the group with a high anchor and the group with a low anchor (or between an anchored

[16] In the example, it does not matter whether the anchor is relevant or not, nor does the anchoring paradigm matter (see section 2.5). One can think of a clearly relevant anchor, for instance the selling price of a similar house in the same street last week, to make the example more understandable.
[17] Cialdini (2007) has demonstrated the power of consistency once people have (publicly) committed themselves to a choice or an opinion.

and an unanchored group) are statistically significant, this is said to be due to anchoring effects. Most often an independent-measures analysis of variance (ANOVA) is used to compare the means of the two groups.

In a set-up with only one anchor, often a high anchor is chosen, but the results can be extended to low anchors. The fact that high anchors create bigger anchoring effects than low anchors is probably explained by a floor effect or "asymmetry of uncertainty" (Jacowitz & Kahneman, 1995, p. 1164). In many problems, e.g. estimating the height of the Mount Everest or the selling price of a house, the target values have a "definite lower bound (zero) but no definite upper bound" (Jacowitz & Kahneman, 1995, p. 1164).

In each of the two measuring approaches the anchor values play an important role. The next section explains how anchor values are usually determined.

2.4.2 Anchor values

Most anchor values are chosen by the experimenters[18], even in situations where the participants might not suspect it. Consider for instance the rigged wheel of fortune (Tversky & Kahneman, 1974) and the loaded dice (Englich et al., 2006)[19]. This raises the question just how the experimenter should determine appropriate anchor values. This process is not standardized in the literature, but it is commonly accepted that setting the anchor values requires a calibration group with unanchored estimates. Jacowitz and Kahneman (1995) suggest using the 15th (85th) percentile of the calibration group as the low (high) anchor. Although they did not explain their choice, the guideline is often followed by other experimenters, who sometimes refer to the 15th and 85th percentile as 'optimal anchor values'. When more than one anchor is used for the same target question, the anchor values are in general determined by a small deviation from the optimal anchor values. Wu et al. (2008), for instance, use a range around the optimal anchor values that does not exceed 5%.

The main contribution of Jacowitz and Kahneman (1995) lies not so much in the specific percentiles they proposed, but rather in the suggestion that anchors could and should be

[18] Sometimes they are not, consider for instance the SSNs used by Ariely et al. (2003). Often the median anchor value of all participants serves as the cut-off between a high anchor and a low anchor group, but Ariely et al. (2003) worked with five groups by using the quintiles as cut-off values.
[19] Englich et al. (2006) had two pair of dice. One pair of dice was loaded to show only 2 and 1, the other to show only 6 and 3. Adding up led to the intended anchor values (3 and 9).

expressed relative to unanchored estimates on the particular target question, rather than specific percentiles they proposed[20]. This makes it possible to refine the concept of plausibility or value alignment.

Anchors can be high or low on a plausible and an implausible level. The optimal anchor values are expected to be plausible – although individual participants may have another perception of the value alignment – because they are based on the distribution of unanchored estimates. To determine a high (low) implausible anchor value, Strack and Mussweiler (1997) added (subtracted) 10 standard deviations to (from) the mean estimate of the calibration group[21].

2.5 Anchoring paradigms

The literature fails to present a clear and collectively exhaustive classification of the experimental design paradigms used to study anchoring effects. These paradigms are not only useful to design experiments, but also to produce anchoring effects in the real world. Before I propose a classification, two different types of anchoring effects are discussed.

2.5.1 Standard and basic anchoring effects

The concept 'standard anchoring effect' gradually emerged to indicate anchoring effects that result from the experimental paradigm introduced by Tversky and Kahneman (1974) in their wheel of fortune experiment. Before participants have to estimate the target value, which is sometimes called the absolute judgement (Wu et al., 2008), they have to complete a comparative judgment task that links the anchor with the target value. Wu et al. (2008) refer to this as a two-stage procedure. For instance, after Tversky and Kahneman (1974) spun the wheel of fortune, they first asked the participants to write down the obtained number and to perform a comparative judgment task – "Is the percentage of African nations among the UN members larger or smaller than the number you just wrote?" (Kahneman, 2011, p. 119). Subsequently they asked the participants to estimate the target value – "What is your best guess of the percentage of African nations in the UN?" (Kahneman, 2011, p. 119). This classic experimental paradigm is now used in the majority of the papers on anchoring effects (Mochon & Frederick,

[20] This implies that a low (high) anchor can be higher (lower) than the actual target value in case the estimates of the unanchored group are far from the target value. In a study of Strack and Mussweiler (1997), the low anchor for the mean winter temperature (in °C) on the Antarctic – determined subtracting one standard deviation from the mean – is -43, even though the actual value is -68 °C. This is also related to Figure 3.

[21] Except when this would lead to strange situations, such as a negative anchor value for a target question about the length of a whale.

2013), which could account for the almost meaningless name 'standard' anchoring effects. Although it is not explicitly mentioned anywhere, it is possible that 'standard' refers to the anchor itself, since Mochon and Frederick (2013, p. 69) state that in this traditional experimental paradigm "participants first judge whether some target quantity is greater or smaller than a presented numeric standard and then render their best point estimate".

More than 20 years later, Wilson et al. (1996, p. 389) discovered the 'basic anchoring effect', which is the influence of an irrelevant anchor in a situation where "people are not asked to consider the anchor as a possible target value". This seemingly trivial difference in experimental design – the absence of a related comparative judgment – has important consequences, which will be discussed after the proposed classification of experimental paradigms.

2.5.2 Classification

The proposed classification hinges on two dimensions (see Table 3). The first dimension corresponds to whether or not a participant is asked to consider the anchor value as a possible target value. The second dimension is the relevance of the anchor, which is determined based on Table 2. It is important to understand that the way in which the anchor is presented, is left out of the classification.

If we interpret the literature strictly, including a comparative judgment[22] (see section 2.5.1), which forces the participant to consider the anchor value as a possible target value, is the only requirement to obtain standard anchoring effects. However, the term is almost exclusively used for irrelevant anchors, as if standard anchoring effects try to mimic the study by Tversky and Kahneman (1974) in every way. To rule out confusion, I propose the classification depicted in Table 3, in which a standard anchoring effect is the influence of an irrelevant anchor in a procedure that includes a comparative judgment. A new expression will be introduced for the influence of relevant anchors in a procedure with a comparative judgment, rather than following the confusing practice of referring to them as standard anchoring effects. Choosing the relevance of the anchor as the second classification dimension, is also supported by the fact that throughout the existing literature the term 'basic anchoring effects' is only used in case of

[22] Throughout section 2.5 'comparative judgment' is short for '*related* comparative judgment'. In section 4.2.3, I will introduce *unrelated* comparative judgments, which do not force the participant to consider the anchor as a possible target value.

irrelevant anchors, even though the use of relevant anchors in a procedure without a comparative judgment is perfectly possible.

A relevant anchor acts as a reference point in the mind of the decision maker[23]. In addition, a comparative adjustment can even force the decision maker to compare this reference point with the target value. Therefore, I propose the expressions 'imposed-reference-point anchoring effects' and 'suggested-reference-point anchoring effects' to describe a design with relevant anchors in a procedure with and without a comparative judgment, respectively. Table 3 gives an overview of the terminology that will be used throughout this dissertation.

Table 3

A classification of anchoring paradigms

		relevance of the anchor	
		relevant	irrelevant
related comparative judgment Is the participant asked to consider the anchor value as a possible target value?	yes	imposed-reference-point AE	standard AE
	no	suggested-reference-point AE	basic AE

Note. AE is an abbreviation for anchoring effect.

2.5.3 Robustness and fragility

Several researchers have demonstrated that standard anchoring effects are a robust phenomenon, meaning that standard anchoring effects have shown a broad scope of impact in numerous experiments (e.g. Ariely et al., 2003; Brewer & Chapman, 2002; Tversky & Kahneman, 1974; Wong & Kwong, 2000). The robustness is at least in part due to the comparative judgment, because it compels the participants to think of the anchor value as a possible target value and because it forces them to pay attention to the anchor (Chapman & Johnson, 2002).

[23] Although irrelevant anchors can subconsciously act as reference points as well, they are not seen as reference points by the decision maker because there is no context alignment and/or no value alignment.

Without a comparative judgment, *relevant* anchors can still produce robust anchoring effects (e.g. Northcraft & Neale, 1987). Based on the demonstrated robustness of standard and suggested-reference-point anchoring effects, imposed-reference-point anchoring effects are expected to be robust as well, since they are characterized by both features that have shown robustness: relevant anchors and a comparative judgment. This is, however, not the case for basic anchoring effects.

A comparative judgment is not essential to produce sizable anchoring effects with irrelevant anchors (e.g. Wilson et al., 1996), but the undisputed robustness of standard anchoring effects stands in stark contrast to what Brewer and Chapman (2002) aptly name the fragility of basic anchoring effects. They illustrated this fragility in their attempt to replicate the study of Wilson et al. (1996) that discovered the basic anchoring effect. Small changes in the experimental set-up, most notably changing the numeric values of the anchor, resulted in the disappearance of the effect (Brewer & Chapman, 2002).

It should be no surprise that merely presenting an irrelevant anchor is often not enough to produce a sizable basic anchoring effect. Otherwise every number could act as an "incidental environmental anchor"[24] and influence judgments. Although the applicability of basic anchoring effects is somewhat more limited, not all basic anchoring effects are as fragile as has been suggested in the years after the turn of the century (Critcher & Gilovich, 2008). Despite the effort of many researchers, defining the additional requirements to obtain sizable basic anchoring effects remains an interesting research topic.

Wilson et al. (1996) suggest that the size of the anchoring effect is positively correlated with the level of attention paid to the anchor[25]. A higher level of attention can be obtained by presenting the anchor in a salient way – for instance showing it repeatedly or using a red color and bold font style (Wu et al., 2008) – but also by forcing the participants to perform some mental processing of the anchor value. Copying the same anchor value multiple times worked for the numeric values used in Wilson et al. (1996), but could not be reproduced by Brewer and Chapman (2002) with other numeric values. Another option, supported by both Wilson et al. and Brewer and Chapman, is extensive processing in the form of a computation task. One

[24] An expression coined by Critcher and Gilovich (2008).
[25] This is considered to be true for basic anchoring effects, but is proven to be wrong for standard anchoring effects by Wu et al. (2008).

example is writing down the three numbers before (4424, 4423, 4422) and after (4426, 4427, 4428) the anchor value (in this case 4425). Although including these types of computation do produce sizable anchoring effects, they will at the same time create too much suspicion and annoyance to be used in a business application. In section 4.3, I will propose a more suited way of integrating computation in a basic anchoring paradigm.

Subliminal anchoring, a more recently discovered experimental design for basic anchoring effects[26], deals with salience and mental processing in a completely different way (see Chapter 1). Mussweiler and Englich (2005) asked participants to look at a computer screen and focus on a fixation point, in this case a changing letter string, for 60 seconds. During this minute, an anchor value was presented ten times, but only for 33 milliseconds[27]. Due to the persistence of vision – the phenomenon that also accounts for the blending of images on a thaumatrope – the numbers are not consciously perceived, and the participants are not even aware that they have seen a number. Nevertheless, their judgment of the target value has been influenced by the anchoring effect.

2.5.4 Debiasing

Many studies show that debiasing is particularly difficult for standard and basic anchoring effects (e.g. Tversky & Kahneman, 1974; Wilson et al., 1996). Wilson et al. (1996) found anchoring effects even after they forewarned the participants by exactly explaining how the anchoring effect could impact their judgment. Wilson and Brekke (1994) propose four criteria that need to be fulfilled to be able to debias a judgment. The first three are interesting when examining anchoring effects.

First of all, the decision maker must be aware of the bias. This condition is already difficult to fulfill. Consider for instance how the experts in the experiment by Northcraft & Neale (1987) denied the influence of the listing price. Being aware of the bias is especially difficult for basic anchoring effects since the anchors are perceived to be irrelevant, and there is no comparative

[26] Technically, subliminal anchoring can be used for every type of anchoring effects in Table 3, but it would we strange to include a comparative question in a situation where the anchor is not consciously perceived by the participant. Additionally, one could question whether context alignment is possible if the participant is not aware of the anchor. Therefore, the subliminal anchoring paradigm is, in my opinion, an experimental set-up for basic anchoring effects.

[27] These are the details of Mussweiler and Englich's (2005) Study 2. Slightly different choices of the fixation point, the sequence of the frames, and the duration of the frames, also led to sizable anchoring effects, which demonstrates that the results of the subliminal anchoring paradigm are quite robust.

judgment that could act as a red flag to the decision maker. Moreover, in the case of subliminal anchoring it is simply not possible that the participant is aware of the bias.

Secondly, the decision maker has to be motivated to debias his judgment. Tversky and Kahneman (1974) showed in their wheel of fortune experiment that is not a sufficient condition for debiasing. They demonstrated that offering pay-offs for the accuracy of the provided estimate did not reduce the anchoring effect. In an experiment different from the SSN experiment Ariely et al. (2003) also proved that providing incentives did not reduce the anchoring effect.

The third condition is probably the most problematic: the decision maker needs to be aware of the direction and the magnitude of the anchoring effect. Once the decision maker has identified the anchor value, the direction of the bias can easily be determined. However, people have difficulties to determine the magnitude of biases. In the context of anchoring effects, this is not surprising since the phenomenon is largely built upon the idea of 'insufficient' adjustment.

In summary, the three conditions are often not fulfilled. Therefore, anchoring effects are especially difficult to debias.

2.6 Conclusion

In this chapter, I provided a critical review of the highlights of the anchoring literature. Many of the oldest themes in the literature on anchoring effects are still awaiting their undisputed conclusions, but at the same time, Epley and Gilovich (2010) suggest that the current literature should make a substantial leap and "unbind anchoring from the dominant experimental paradigm that has been used to study it" (p. 21). The present master dissertation answers their call and proposes a new way of presenting the anchor values (see section 4.3), but also partly relies on the demonstrations of the basic anchoring by Wilson et al. (1996) and the traditional anchoring paradigm, which produces robust standard anchoring effects, that was invented by Tversky and Kahneman (1974).

3 Price judgments

Price judgments encompass willingness to pay (WTP) or willingness to accept (WTA) judgments, and valuation tasks. The effect of anchoring on these price judgments has been studied widely, but almost exclusively for relevant anchors, which are called reference prices. These reference prices can be used to sell more and/or at a higher price. Selling more can be achieved by the use of suggested retail prices. This is probably the most famous practical illustration of the anchoring effect and will therefore be discussed in section 3.2.2. Based on a study by Northcraft and Neale (1987) I explained in chapter 1 how anchoring effects can raise the appraised value of a house, which implies that it can be sold at a higher price.

Standard anchoring effects – effect that result from the use of irrelevant anchors in a procedure with a comparative judgment (see Table 3) – on price judgments are also covered in the literature, but clearly less often than reference prices. Most of the time the price judgments are simple and hypothetical valuation tasks such as estimating the price of a midsize car (Mussweiler & Englich, 2005). Nevertheless, some important research has been done on WTP and WTA judgments (e.g. Ariely et al., 2003; Green, Jacowitz, Kahneman, & McFadden, 1998; Simonson & Drolet, 2004).

The papers listed in this section are important, because they provide a basis to study *basic* anchoring effects on price judgments, which has barely been done in the existing literature. A notable exception is Wu et al. (2008), who demonstrate that basic anchoring effects can influence WTP. Unfortunately, their results are not conclusive (section 4.3), which caused me to try to demonstrate that basic anchoring effects can indeed influence WTP. But first it is necessary to discuss how consumers evaluate transactions, and the difficulties that arise when researchers start to look into this.

3.1 Purchase decisions

Consider the case of Mr. Econ and Ms. Human who find themselves looking at the shop window of a smartphone store. They both agree that this particular shop window is an architectural masterpiece, but they are particularly interested in a smartphone with a price of €400 instead of the suggested retail price of €500. Should they buy this smartphone?

Mr. Econ[28] makes purchase decisions according to rational choice theory, the standard economic view on consumer choice. He is rational, logically consistent, and able to make optimal decisions. Mr. Econ is said to focus on self-interest and therefore only capable of instrumental behavior. In reality, it does not make sense to study purchase behavior assuming that everyone acts exactly like Mr. Econ. Therefore, Ms. Human's behavior will illustrate a more realistic dynamic behind purchase decisions.

3.1.1 Rational choice

According to rational choice theory, Mr. Econ should compare the utility of buying and owning the smartphone with its opportunity cost. These two elements are dependent on the individual under observation, as is the final outcome of the purchase decision. Utility expresses the happiness or value arising from an experience and is an attempt to compile the preferences of an individual into a one-dimensional construct. The opportunity cost is economic jargon for the value of the best alternative; in our example the opportunity foregone by buying the smartphone. If Mr. Econ does not purchase the smartphone, what can he do with the resources otherwise invested in the decision? What can he do with €400 and the thirty minutes that he does not spend in the store to buy the smartphone? And the most important question: will the best of all possible alternatives provide a higher utility to Mr. Econ than the utility of buying and owning the smartphone?

Suppose Mr. Econ likes the purchase more than his best alternative, in this case inviting a longtime colleague for a €400 thirty-minute lunch[29]. In that case, the utility of buying and owning the smartphone is bigger than the opportunity cost. Mr. Econ buys the smartphone.

The out-of-pocket cost or financial outlay – in this case the price of the smartphone – is the most prominent part of the opportunity cost (Thaler, 2015)[30,31]. If the price of the smartphone

[28] Richard Thaler, the 2017 winner of the Nobel Memorial Prize in Economic Sciences, introduced 'Econ' – as opposed to 'Human' – instead of the rather old-fashioned expression 'homo economicus' (Kahneman, 2011; Thaler, 2015, p. 4).
[29] After all, Econs are said to be selfish. Also notice that Mr. Econ has plentiful money, otherwise his best alternative for the money (€400) would probably not be a thirty-minute dinner.
[30] Thaler (1985) and other researchers sometimes define opportunity costs more narrowly, i.e. without the out-of-pocket cost.
[31] This does not mean that the out-of-pocket cost always constitutes the main part of the opportunity cost. If you receive a gift, the financial outlay or out-of-pocket cost is zero, but there still is a substantial opportunity cost because you can sell the gift to someone else. In fact, if the gift can be sold at a price higher than the receiver's reservation price, no person acting according to rational choice theory will be happy with a gift he did not want to buy with his own money (Thaler, 1985).

increases, Mr. Econ faces a growing opportunity cost. At a certain point the price will reach a level where the opportunity cost offsets the utility. At this level, Mr. Econ is indifferent between buying and not buying the smartphone. Beyond this level, he will not buy. This specific price is referred to as Mr. Econ's reservation price or willingness to pay. Buying at a lower price than the reservation price will produce some excess utility for Mr. Econ, which is called consumer surplus.

In the example above, Mr. Econ will end up buying the smartphone. In fact, it is not only what he will do, it is also what every reasonable person should do in his case. Comparing utility and opportunity cost is a so-called normative theory, a guideline for optimal and logically consistent decisions. Normative theories are extremely useful if interpreted in this way. In standard economics, however, these theories have also been used to describe human behavior, which is incorrect because most humans – even economists – do not make decisions like Mr. Econ.

Standard economic theory assumes that preferences are complete, reflexive, and transitive (Varian, 2014). However, even if these assumptions hold – which is often not the case[32] – another more fundamental problem arises: decision makers do not know their preferences and utilities. In addition, it is almost impossible to be fully informed on all foregone opportunities. Even if there is complete information on all alternatives it is likely that correctly processing the information would be too difficult and would take a lot of time[33]. To make numerous decisions each day would mean there is no more time left for living (Thaler, 2015).

A small but significant tweak to the normative theory can blast these arguments away. After all, people are able to compare the *expected* utility with the *expected* opportunity cost. Due to everything mentioned above, these expectations are likely to differ from the actual values. Depending on the complexity of the purchase decision and the ability of the decision maker to form his expectations, the outcome will be slightly or severely worse than achieved by acting according to the normative theory. Allowing decision making based on expectations is equivalent to gradually abandoning the normative character of the theory[34].

[32] Consider for instance preference reversals due to framing, which is aptly described in Tversky and Kahneman's (1981) Asian disease problem.
[33] The time needed to gather and process information could be taken into account when calculating the opportunity cost, but this introduces a new level of complexity since this requires gathering and processing information with regard to the time needed to gather and process information. Technically, this recursion could keep going.
[34] In this and the previous paragraph, I implicitly referred to bounded rationality, intended rationality, and satisficing. For a more adequate overview see Simon (1955), Jones (2002), and Simon (1979), respectively.

But there is another reason why a normative theory is unable to accurately describe actual behavior. Humans are influenced by factors that standard economic theory believes to be irrelevant (Thaler, 2015). Why do people buy things they never use? Why do people refrain from making obviously good decisions like saving for retirement? Standard economic theory has no answers. In the imaginary world of Mr. Econ bad decisions do not exist and good decisions are always made.

3.1.2 Human behavior

Thaler (1985) proposes a descriptive or positive theory to explain and predict real consumer behavior. His theory builds upon the normative theory, but distinguishes two different types of utility.

The acquisition utility captures what standard economic theory calls consumer surplus or deficit (Thaler, 2015). It is defined as the difference between the utility or value of the good received and the opportunity cost, which Thaler simply calls the (financial) outlay – probably because people tend to give too much weight to the out-of-pocket component of the total opportunity cost. If the story ended with the assessment of acquisition utility, the only tools a company has to influence purchase decisions are price and product characteristics. Marketing elements related to framing (Tversky & Kahneman, 1981) have no effect on the acquisition utility (Thaler, 1985).

Transaction utility, for its part, can be defined as the difference between the price paid and what you expected to pay (Thaler, 1985). In what follows, this expected price is called the internal reference price (IRP) to distinguish it from reference prices, which are relevant anchors in a context of price judgments. The transaction utility captures the "perceived quality of the deal" (Thaler, 2015, p. 59), something Ms. Human cares about – unlike Mr. Econ. If the price Ms. Human has to pay is lower than what she expected or finds just, she will think of the deal as a "bargain"[35] (Thaler, 2015, p. 59). If instead the price exceeds her IRP, she will begin to think of it as a "rip-off". Notice that the height of the price paid hardly matters. It is Ms. Human's IRP that determines whether she perceives a high price as an "expected annoyance" or a rip-off (Thaler, 1985, p. 207).

[35] Mr. Econ can also encounter a bargain, but this is purely based on a positive acquisition utility or consumer surplus.

A positive (negative) transaction utility can offset a negative (positive) acquisition utility. Therefore, Thaler's transaction utility theory is able to explain and describe actual human behavior, like buying things we do not need only because we consider them to be cheap or failing to make decisions that would have resulted in a positive acquisition utility. Ms. Human might decide not to buy the smartphone although this would give her a positive acquisition utility, simply because she expected the smartphone to cost much less than €400. In our smartphone example the reverse is more likely. Even though she does not need a smartphone, she might buy it only because it is sold at a discount and therefore poses an irresistible deal.

3.2 Willingness to pay

Given Thaler's transaction utility theory, the concept of willingness to pay should be broadened. In standard economics, the price where the (acquisition) utility equals the total opportunity cost is referred to as the reservation price. In the context of the present dissertation, the willingness to pay (WTP) is used to describe actual human behavior and therefore no synonym of the reservation price. In what follows, the WTP or highest acceptable price of an individual is the highest price where the combination of acquisition and transaction utility is still positive.

3.2.1 Observing willingness to pay

Observing the exact WTP of a consumer is extremely difficult because it is not revealed by accepting or rejecting a transaction, except in certain auctions or negotiations. Thaler (1985) distinguishes two stages in each purchase process: an evaluation of the transaction and a decision to accept or reject it. Since then, research has supported Thaler's view on the purchase process (Monroe, 2003). For most transactions, consumers do not make the effort in the evaluation stage to consciously determine their own WTP based on the acquisition and transaction utility. They do not ask themselves 'At which price would I buy?', but rather 'Should I buy or not at this particular price?'[36].

Just asking participants to reveal their own WTP will not work, not only because the participants are in a cold state and there is nothing at stake, but also because their WTP is not known to them as well. According to the literature on 'constructed preferences', it is only during the decision-making process that the preferences – and the WTP – are 'constructed' and

[36] This is an example of Kahneman's (2011) view on heuristics: a complex target question is substituted for a simpler question, the heuristic question.

'imprinted' (Ariely et al., 2003; Slovic, 1995). Choice-based conjoint analysis can be used to help consumers to discover and reveal their WTP. However, both 'just asking' and choice-based conjoint analysis require consumers to consider a purely hypothetical situation. Furthermore, without real consequences, there is no way to verify the obtained WTPs.

Researchers like Ariely et al. (2003) and Simonson and Drolet (2004) therefore conduct experiments in settings with real money, real products, and real consequences in order to create a situation where it is in the participant's best interest to reveal his WTP. To achieve this incentive-compatibility, Ariely et al. (2003) use the Becker-DeGroot-Marschak method (Becker, DeGroot, & Marschak, 1963). After a participant has communicated his WTP, a computer draws a – to some degree random – price. If this price is lower than the WTP, the participant has to pay the drawn price and receives the product. If not, the participant is not granted the opportunity to buy the product, not even at the drawn price. Another procedure is the Vickrey auction[37]. In this type of sealed-bid auction the product is sold to the highest bidder, but at a price equal to the bid of the second-highest bidder (Vickrey, 1961).

3.2.2 Influencing willingness to pay

Based on Thaler's transaction utility theory, a seller has several tools to increase revenue at a given price, i.e. to sell more. Offering a better or better wanted product is one of the tools described in standard economics[38]. This increases the value for the buyer (the acquisition utility) and therefore attracts more buyers. However, the possibility to manipulate the internal reference price (the transaction utility) provides a whole new toolbox to sellers. In the same article where he presents the transaction utility theory, Thaler (1985) outlines two categories of tools to influence the IRP: obscuring and increasing the perceived reference price.

Obscuring comes in handy when the selling price of the product is higher than the consumer's IRP. An example is selling unusual sizes or formats, so that consumers have no clearly established IRP and the transaction disutility becomes less obvious (Thaler, 1985).

[37] The method is named after William Vickrey, who was the first author to describe it. He won the Nobel Memorial Prize in Economic Sciences in 1996, three days before he passed away.
[38] In standard economics, the most important way to increase revenue is a price change, all the while taking into account the price elasticity of demand. In this paragraph price changes are left out of the discussion.

To increase the IRP, a seller can create the perception that he faced more costs to produce or sell the product, even without actually altering the product or the experience for the consumer (Thaler, 1985)[39]. Fairness is an important element in establishing an IRP. The higher the perceived costs for the seller, the more likely it becomes that a consumer will expect and tolerate a higher price. In turn, because the IRP is higher, the transaction utility from the purchase has increased. Another tool to increase the IRP is the use of suggested retail prices (SRPs) (Thaler, 1985), also known as manufacturer's suggested retail prices or recommended retail prices.

Back to Mr. Econ and Ms. Human. The fact that the price of the smartphone (€ 400) is lower than the original price or SRP (€500), rather than priced at €400 from the start, makes no difference to Mr. Econ. Neither does the fact that the shop window is a masterpiece of design. For Ms. Human, however, the shop window is an indication that the smartphone seller faced more costs, which might increase her IRP. The SRP of €500 does manipulate her IRP and tricks her in to believing the smartphone is a bargain. In addition, the SRP might also serve as an indication of the quality of the product, which means that the perceived value of the smartphone (the acquisition utility) can be influenced as well (Thaler, 2015).

If a seller were able to manipulate each consumer's IRP individually, this would be in some ways the opposite of what is known as perfect or first-degree price discrimination. In a situation of perfect price discrimination, a monopolist, not willing to produce more than a certain level, charges each individual his or her WTP – as long as this WTP lies above the seller's marginal costs – transferring the complete consumer surplus to the seller (Varian, 2014). In the case of IRP manipulation, a firm can refrain from charging every consumer his or her WTP and instead try to elevate the WTP of every consumer until it reaches at least the price the seller asks for the product.

Note that any manipulation of the IRP can increase the quantity demanded at a given price in two ways, similar to the so-called two laws of demand that describe the effect of a price reduction in standard economic theory. The first mechanism is based on an increase of the total number of buyers because – thanks to the higher transaction utility – people will find it

[39] If this practice can be performed better by a firm than its competitors, this means that even firms in perfectly competitive markets have the opportunity to generate more demand than their competitors, which enables the firms to outperform their competitors due to (positive) internal economies of scale.

worthwhile to buy the product. According to the second law, the number of items per buyer will increase as well. Higher demand at the same price, and its equivalent, a higher price without a drop of demand, imply higher revenues for a seller who can successfully manipulate the IRP of his consumers.

3.3 Conclusion

Consumers do not make purchase decisions according to rational choice theory. Their susceptibility to the perceived quality of the deal provides new tools to sellers. A seller who successfully raises the internal reference price (IRP) of his consumers, has at the same time raised his revenue. Suggested retail prices, which produce a suggested-reference-point anchoring effect (see Table 3), are widely used to manipulate consumers. In chapter 4, I propose a procedure to influence the IRP based on irrelevant anchors in general, and basic anchoring effects in particular.

4 Practical application

This chapter combines the insights of the previous chapters on anchoring and price judgments to outline an application that is based on basic anchoring effects and influences the willingness to pay of consumers. At the end of this chapter, in section 4.3, I explain why hiding the anchor values in CAPTCHAs can provide the key to the intended business application. But first the choice for an e-commerce setting is justified. Subsequently, the research questions are presented and discussed. These questions are related to the search for a practical application, but will also contribute to the anchoring literature.

4.1 E-commerce setting

Peterson and Merino (2003) believe that consumers, and especially Internet shoppers, have more power nowadays because the Internet[40] gives them access to enormous amounts of data. This enables them to compare numerous offers and look at feedback provided by other consumers, both at virtually no cost. However, there are several reasons to question a shift of power in favor of consumers. The advantages of the information abundance are limited due to the bounded rationality of consumers (Simon, 1955), who will more often resort to heuristics, which potentially leads to biased judgments. Furthermore, access to more data does not guarantee an increase in the quality of the data on which consumers base their decision. In addition, web design can be used to influence consumers (Mandel & Johnson, 2002). Hiding anchors to provoke basic anchoring effects is relatively easy and inexpensive, and even subliminal anchoring becomes possible.

The increased use of heuristics in a complex online world and the opportunity to integrate anchors in website design, make an e-commerce setting the perfect start in the search for a practical application involving (basic) anchoring effects. Unlike Wu et al. (2008), I will examine so-called fixed price-based transactions in which a consumer accepts or rejects the price determined by the seller. In the world of e-commerce, these so-called fixed price-based transactions are more common than auction-based transactions. Besides, studying auctions requires the simulation of an English auction, which is more difficult than simulating a sealed-bid auction like the Vickrey auction.

[40] When Peterson and Merino (2003) mention ‘Internet’ they actually refer to the World Wide Web (which can be accessed via the Internet).

4.2 Research questions

In search of a practical application for basic anchoring effects, the existing theories and frameworks are tested. Four main research questions are set out below, all related to the same idea: using anchors to influence the consumer's internal reference price in order to increase his willingness to pay.

4.2.1 Q1: Can basic anchoring effects influence willingness to pay in an e-commerce setting?

The first research question (Q1) lies at the heart of the search for a practical application involving basic anchoring effects. If basic anchoring effects can be used to raise the internal reference price of Internet shoppers, this can eventually lift the WTP of consumers and the revenue of the seller (see section 3.2.2). It is worthwhile to examine the possibility of using basic anchoring effects to obtain this, even though standard anchoring effects and reference-point anchoring effects have shown their robustness and power in numerous experiments.

Using irrelevant anchors, rather than relevant anchors, may seem odd. However, the literature on reference prices is extensive and their workings widely known. Unfortunately, companies are more acquainted with the benefits of reference prices than consumers are familiar with the pitfalls. Furthermore, irrelevant anchors are definitely not less practicable than reference prices. Business sense indicates limitations as to the number and type of products to which sellers can assign reference prices[41]. Besides, the numeric values that a seller can use for these reference prices are constrained by those of his competitors. Conversely, using irrelevant anchors does not rule out target products or numeric values and even allows attaching multiple anchors to a product.

The choice for a procedure *without* a comparative judgment is more straightforward to justify than that for irrelevant anchors. Explicitly asking consumers to compare a numeric value with their expected price or WTP would be ridiculous in practice. Consumers would perceive the comparative question not only as annoying, but also as suspicious. The latter also limits the available numeric values since using implausible anchors would increase these feelings of suspicion.

[41] Thaler (2015) states that using suggested retail prices works best for products that are purchased infrequently and products whose quality is difficult to assess.

Even without the arguments in favor of using irrelevant anchors in a one-stage procedure, studying basic anchoring effects can be justified by the potential contribution to the existing literature. Basic anchoring effects are expected to produce significant increases in the price estimates (e.g. Brewer & Chapman, 2002; Wilson et al., 1996; Wu et al., 2008), but this depends on the specific experimental paradigm that is used (see section 2.5.3). In section 4.3 an anchoring paradigm suited for the creation of basic anchoring effects in an e-commerce setting will be presented.

4.2.2 Q2: Are basic anchoring effects less effective than standard anchoring effects in influencing WTP in an e-commerce setting?

Although the basic anchoring effect is a good candidate to influence WTP, its effectiveness needs to be demonstrated. The existing scientific literature claims that standard anchoring effects will be larger and therefore more effective than basic anchoring effects (e.g. Ariely et al., 2003; Simonson & Drolet, 2004; Wu et al., 2008). Additionally, basic anchoring effects are considered to be fragile (Brewer & Chapman, 2002). Comparing the results obtained by using standard anchoring with those of basic anchoring effects will shed light on the effectiveness of basic anchoring effects. If the effectiveness of the standard effect turns out to be bigger, a trade-off for the practical application emerges: working with a low-annoyance, low-suspicion, but small basic anchoring effect or a more annoying and suspicious, but also more impactful standard anchoring effect.

4.2.3 Q3: Does an experimental set-up with a comparative judgment that is not related to the target value produce anchoring effects?

The third research question (Q3) is predominantly of academic value. In the standard experimental paradigm, the experimenter asks the participant to compare the anchor value with the target value. In the wheel of fortune experiment both the comparative and the absolute judgment included the target, in this case the percentage of African nations in the UN (Tversky & Kahneman, 1974). Instead of relating the comparative judgment to the target value, Tversky and Kahneman could have chosen an *unrelated* comparative judgment. An example of a factual question that is not related to the target value would be: Is the percentage of students at your university who fail to pass their Microeconomics exam larger or smaller than [anchor value]?

Based on the classification in Table 3, comparing an irrelevant anchor value to the answer on a question other than the target question still leads to a basic anchoring effect. However, if the inclusion of an unrelated comparative judgment produces anchoring effects that are similar to *standard* anchoring effects, refining the classification might be necessary. However, one expects to see smaller effects than those observed in the standard experimental paradigm because the unrelated comparative judgment does not force the decision maker to consider the anchor as a potential target value.

Besides the academic relevance of this research question, there could be a practical application for the unrelated comparative judgment as well. As stated in section 2.5.3, suspicion is one of the disadvantages of including a comparative judgment in a set-up to influence price judgments in an e-commerce setting. An *unrelated* comparative judgment could reduce this suspicion because the link with the price judgment is hidden[42]. Unlike with a related comparative question, there is no restriction on the possible anchor values when an unrelated comparative question is used, simply because there is no plausibility check with regard to the target value and therefore no opportunity for implausibility to increase suspicion. One would also expect multiple anchors, and hence multiple comparative questions, to be less suspicious when there is no relation with the price judgment. In conclusion, a comparative judgment that is not related to the target value could be a better procedure to influence price judgments than the standard experimental paradigm provided that the resulting anchoring effect is not too small. Once again, a trade-off might emerge between size on the one hand, and suspicion and annoyance on the other.

4.2.4 Q4: Do standard and basic anchoring effects vanish when confronted with suggested-reference-point anchoring effects?

Examining a practical application for basic anchoring effects requires an experiment with sufficient external validity. In reality, Internet shoppers have access to many websites. Although the seller can try to obscure the IRP (see section 3.2.2), consumers can often find reference prices, which act as relevant anchors and produce suggested-reference-point anchoring effects (see Table 3). It is therefore important to investigate whether basic and standard anchoring effects vanish if consumers encounter a reference price.

[42] Conversely, one could argue that an unrelated question, which appears seemingly out of nowhere, creates more suspicion. However, in section 5.4 I propose a way to include an unrelated question that is less suspicious.

Based on the anchoring and adjustment account, one would expect that the reference price will make consumers adjust away from their initial IRP – which was subject to an anchoring effect – towards the reference price (Tversky & Kahneman, 1974). However, one expects that this adjustment will be insufficient to eliminate the initial anchoring effect (Wilson et al., 1996).

In their SSN experiment, Ariely et al. (2003) demonstrated "coherent arbitrariness". After the students wrote down their bids for the bottle of wine, Ariely et al. put several other products up for auction. Although the bid for the bottle of wine was 'arbitrary' due to the anchoring effect, the bids for the other products were 'coherent' with this first bid. Every student provided for instance a lower bid for the box of chocolates than for the bottle of wine. However, due to the manipulated first bid, the students with a high-ending SSN wanted to pay more on average for the box of chocolates than the students with a low-ending SSN ($21 and $10, respectively).

In a way, the hypothesis related to Q4 can also be described by the idea behind coherent arbitrariness. The standard and basic anchoring effects will manipulate the arbitrary initial IRP. After this IRP has been imprinted, coherence ensures that the subsequent suggested-reference-point anchoring effect does not wipe out the standard and basic anchoring effects.

4.3 Integrating anchors in website design

The choice for a procedure with or without a (related) comparative judgment, is an important element of the presentation, but it provides no guidance as to when and how the anchor should be presented, not even in a procedure with a comparative judgment. It would seem straightforward to introduce the anchor as part of the comparative question, but Tversky and Kahneman (1974) deliberately used a wheel of fortune to present the anchor right before the comparative question. Although nowhere explicitly mentioned, this is done to show the participants the arbitrariness[43] or absurdity of the number they are later asked to compare with the target value. Besides the perception of arbitrariness, suspicion, attention, and mental processing should be taken into account when designing a practical application. An additional element in the context of an academic paper is scientific rigor.

[43] In reality the numbers obtained by spinning the wheel of fortune were not arbitrary because the wheel was rigged. Ariely et al. (2003) did use truly random anchors in their SSN experiment and grouped the participants based on the the quintiles of the SSN distribution. Tversky and Kahneman (1974) could have used a similar procedure instead of sabotaging the wheel of fortune.

Two studies have come up with ideas on how to present anchors in a subtle way in the case of basic anchoring. Wu et al. (2008) suggest putting the irrelevant anchors in the product description, and Critcher and Gilovich (2008) experiment with anchors in the product name of a cellphone or restaurant. However, neither paper displays impeccable scientific rigor. By using the product name, Critcher and Gilovich (2008) might unintentionally provide a clue with regard to the quality of the product. Many consumers would prefer the P97 smartphone to the P17 even when the two phones are identical in every aspect. Wu et al. (2008) incorporate the anchor values in the product description as the fictional number of professionals who have tested and approved the smartphone, the fictional number of people who have purchased it in advance, and the fictional number of sales offices that are selling the product. By doing this, Wu et al. (2008) add other psychological phenomena, such as social proof, to the mix. Once again, many consumers would prefer a smartphone approved by more professionals, purchased by more people, and sold in more sales offices. A quality hint, social proof, and other psychological phenomena are good tools for a practical application, but accidentally including them in an experiment harms its scientific rigor. Although both studies observed a higher willingness to pay in the high anchor conditions, there is no way to determine what share of this increase is due to the anchoring effect.

I want to introduce a new, effective, and practicable way of integrating anchors in website design: presenting anchors in CAPTCHAs. A CAPTCHA or 'Completely Automated Public Turing test to tell Computers and Humans Apart' is widely used by websites to verify that they are dealing with human users. To pass the test, a person typically has to distinguish and copy a combination of letters and numbers from a distorted picture. Figure 4 shows how an anchor value can be presented to Internet shoppers.

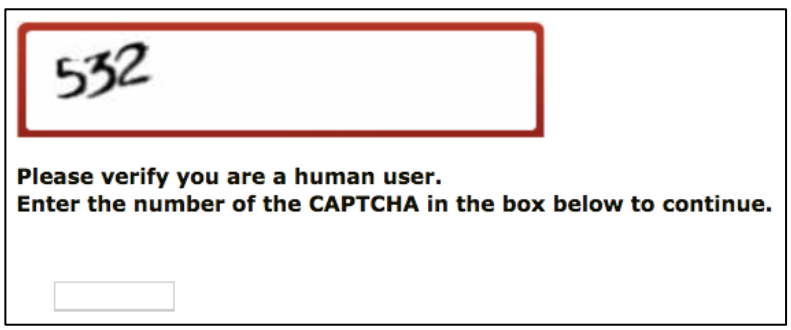


Figure 4. A CAPTCHA introducing the anchor value, in this case 532.

The anchor value is a round number – 532 instead of 532.99 – to avoid a possible dual anchor effects caused by the number before the decimal sign (532) and by the decimals (99). Also, 532.99 is expected to raise more suspicion than 532 because it looks less like an arbitrary combination of numbers and more like a price. This said, the choice for a traditional CAPTCHA has the important repercussion that there is no way to be sure that people actively think of the number as one (five hundred thirty-two) rather than as three separate numbers (five-three-two). A possible solution to this will be presented in section 5.4.

The CAPTCHA set-up has four substantial advantages. The first advantage is that a CAPTCHA does not raise suspicion, and therefore limits the perception of manipulation. One can also assume that the consumers think that the symbols of the CAPTCHA are arbitrarily chosen by a computer, rather than by the seller. A second advantage is that it allows the seller to take control over the situation. The CAPTCHA forces the participant or consumer to pay attention to the anchor, which is paramount to effectively produce basic anchoring effects[44]. In the existing literature on basic anchoring effects, experimenters often have difficulties to ensure that every participant actively looks at the anchor. A solution used by Wu et al. (2008) was to use a red bold font style for every anchor value, but this seems hardly practicable in a business application based on irrelevant anchors. Using CAPTCHAs seems to allow to escape the trade-off between limiting suspicion and ensuring a sufficient level of attention.

The ability to impose minimal mental processing (see section 2.5.3) of the anchor value without causing suspicion is a third plus. After all, copying the anchor value is a standard feature of CAPTCHAs. Using CAPTCHAs also enables the seller to force the consumer to do more advanced mental processing under the guise of trying to 'tell computers and humans apart'. One example is asking the consumers to compare two CAPTCHA values and to copy the largest of the two numbers (see Figure 5)[45]. Moreover, Mead and Hardesty (2018) demonstrate that using disfluent fonts already leads to bigger anchoring effects since the disfluency attracts more attention and reading these fonts requires more mental effort. Finally, the proposed CAPTCHA set-up also ensures scientific rigor, in the sense that the basic anchoring effect is captured without contamination by other psychological phenomena.

[44] Except in the subliminal anchoring paradigm (see section 2.5.3).
[45] Another option is asking the consumers to add up two numeric CAPTCHAs. This is, however, less desirable since the sum can act as a third higher anchor value.

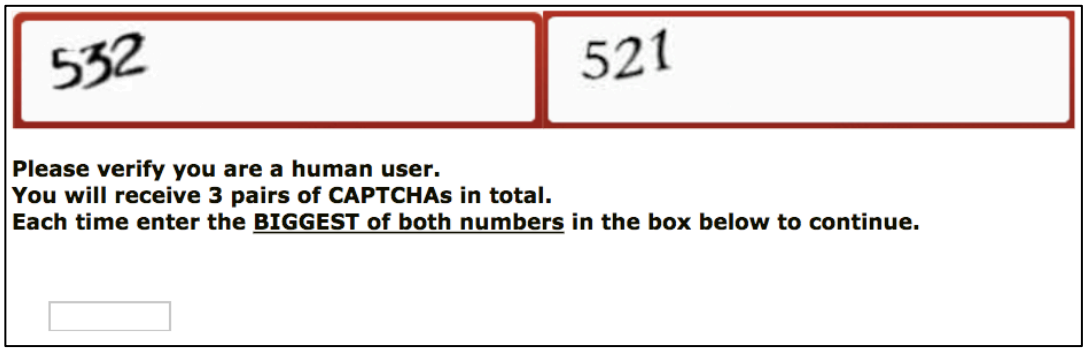


Figure 5. One of the three CAPTCHAs used in the 'computation condition'.

Now that we have a way to integrate anchor values in website design, the research questions related to the practical application have to be answered. In chapter 5, the conducted experiment will be discussed. Chapter 6 will thereafter present the results of this experiment and the answers to all research questions.

5 Experiment design

Chapter 5 discusses the methodology of the conducted experiment. First, the choice for a smartphone as target product is explained in section 5.1. In section 3.2.1, I highlighted the difficulties that arise when an experimenter tries to measure willingness to pay. Based on Thaler's transaction utility theory, section 5.2 will provide an answer as to how this experiment will measure WTP. Section 5.3 and 5.4 present the used anchor values and the experimental conditions, respectively. At the end of this chapter, right before discussing the results, we will have a look at the chosen population and the used sampling method.

5.1 Target product

Previous studies on the effect of anchoring on price judgments have used a wide variety of target products. Ariely et al. (2003) use, among others, a cordless trackball, a cordless keyboard, wine, and Belgian chocolates. Simonson and Drolet (2004) use a 'cordless phone' and, some years later, a more modern mobile phone acted as the target product in studies by Wu et al. (2008) and Wu, Cheng, and Yen (2012). Most target products are high-end products of an unknown brand. This minimizes product familiarity, which is a moderating factor inversely related with the size of the anchoring effect (Van Exel, Brouwer, van den Berg, & Koopmanschap, 2006). To build on the existing studies with target products that allow obtaining sizable anchoring effects, a smartphone is used as target product.

The specific smartphone – see Figure 6 – was chosen for a couple of reasons. First of all, the brand and model of the phone were determined to minimize the probability that participants recognize the smartphone and are familiar with its real price[46]. The dual camera – not a feature of common smartphones – also aims to reduce product familiarity. On the other hand, the technical specifications of the smartphone are particularly good, which increases the perception of a high-end product. It is a potential drawback that the product description contains a lot of numbers, because these could in theory act as anchors. The specifications were nonetheless included because the participants need some information to establish their price estimates. The

[46] The brand and model of the smartphone are not mentioned in the experiment. The participants received a picture of the smartphone from which all logos were removed, together with the actual specifications of the smartphone. It hardly matters in the context of this experiment, but the smartphone is a Xiaomi Mi A1 with a retail price around €200 in Belgium at the time of the experiment. However, its specifications are similar to those of smartphones sold at significantly higher prices.

specifications indicate for instance that this specific smartphone is a high-end product, rather than a smartphone priced below €100. In addition, the product description creates a more realistic scenario, which contributes to the external validity of the experiment. It also is important to note that the product description is the same across all experimental conditions (see section 5.4). Since measuring anchoring effects happens in a between-subjects design (see section 2.4.1), the numbers in the product description should not distort the answers on the research questions.

Figure 6. The specific smartphone used in the experiment, together with its specifications.

5.2 Observing anchoring effects

Section 3.2.1 already highlighted the difficulties that arise when trying to measure WTP. In the context of this experiment, 'just asking' participants their WTP is a particularly bad idea. The way participants determine their own WTP can differ, as can their understanding of what WTP actually is. Besides, a participant's WTP might depend on his willingness to buy a smartphone at that specific point in time and on his ability to think of the hypothetical question as a real situation. Too many differences may arise for which the results cannot be controlled, especially with a small sample size. A Vickrey auction or Becker-DeGroot-Marschak procedure is not feasible. With these procedures, the participants should be willing to buy a smartphone, simply because they might end up having to do so. In addition, auctioning off smartphones requires

substantial research budgets. It is no coincidence that fairly cheap products are used in experiments such as those of Ariely et al. (2003).

For these reasons, I will not directly measure the impact of the anchoring effects on participants' WTP, but rather the impact on their IRP by asking each participant their best guess of the price of the target smartphone (as set by the manufacturer).

5.3 Numeric value

Studying the anchoring effects mentioned in Q1, Q2, and Q3 requires irrelevant anchors. Based on the framework in Table 2, three types of irrelevant anchors can be distinguished. In our experiment plausible anchor values will be used to maximize the chances of obtaining sizable anchoring effects (see section 2.3.2). Therefore, the irrelevant anchors in the experiment are 'irrelevant despite plausibility' (see Table 2)[47]. A high anchor is chosen since the goal is to examine whether anchoring effects can raise the IRP, rather than to provoke lower price expectations with low anchors.

Before conducting the actual experiment, a calibration study was set up with 30 business or economics students at the *Vrije Universiteit Brussel* who were between 18 and 25 years old. The participants estimated the target value – the manufacturer's price for the smartphone – without any influence from anchors. Consistent with Jacowitz and Kahneman (1995), the numeric value of the high and 'irrelevant despite plausibility' anchor is determined by the 85th percentile of the distribution of all price estimates from the calibration group. This percentile and anchor value is 532[48]. Certain experimental conditions will require repeating the anchor (see section 5.4). The anchor values used for this repetition are similar but not identical. Using a small deviation of 2%, as is the case in Wu et al. (2008), from the 85th percentile leads to 543 and 521.

Research question four (Q4) requires a relevant anchor or reference price. To achieve this, the context and value of the anchor needs to be related to the target value (see Table 2). Context alignment is assured by stating that the value of the relevant anchor corresponds to the price for a more or less similar smartphone (see Figure 7).

[47] In any description of the experiment or the results 'irrelevant' will be short for 'irrelevant despite plausibility'.
[48] The exact value of the 85th percentile was 532,5. The choice to round down instead of up was made based on a coin toss. This coin toss procedure was repeated for all subsequent values.

Figure 7. The introduction of the reference price.

There is no clear guidance on optimal anchor values for the reference price needed to answer Q4. In the experiment, I chose the median estimate (50th percentile) of the calibration study, which was 367. First of all because this value should be perceived plausible by most participants. The second argument is lengthier. Without any effect of the irrelevant anchor, one would expect the median of all price estimates to be close to the median estimate in the calibration study (367). If the relevant anchor (367) produces an anchoring effect, this will cause half of the participants – the 50% with an initial price estimate below the median – to provide a higher second estimate. Likewise, the other half of the participants will have a lower second estimate. Taking the median of the calibration study as the reference price is therefore a priori the best way to study upward and downward adjustments at the same time. A posteriori, if anchoring effects managed to pull up the first price estimate, there will of course be more participants who adjust downward. From a practical point of view, the focus is also on this downward adjustment, because Q4 tries to find out whether the suggested-reference-point anchoring effect wipe out the standard and basic anchoring effect.

5.4 Experimental conditions

The experiment was conducted via the online survey that can be found in its entirety in Appendix A. The real purpose of the study, influencing WTP with anchoring effects, was hidden and therefore the survey was titled ‘Price estimations in an e-commerce environment’. The first part of the survey tries to resemble a real website experience. First, the participants

encounter a webpage with several smartphones, followed by the target smartphone and a short product description (see Figure 6). Subsequently, the anchor values, if any, are presented, the level of which depends on the experimental condition. To address all research questions, five experimental conditions are required. They differ in terms of the presentation of the anchor and the comparative judgment (see Table 4 for an overview).

Table 4

Main differences between the experimental conditions

condition	CAPTCHA			comparative judgment
	type	number	instruction	
no anchor	letters *eyr*	1	copy	/
standard anchoring effects	number *532*	1	copy	related *smartphone price*
unrelated comparison	number *532*	1	copy	unrelated *height Atomium*
copy	number *543, 521, 532*	3	copy	/
computation	number *543, 521, 532*	3 times 2	copy highest number	/

The first condition is the ‘no anchor condition’, which will be used as a basis for comparison. I have chosen to include an unanchored condition in the main experiment even though it was in principle possible to use the results of the calibration group (see section 5.3). This ensures that the unanchored participants are even more similar to the participants in other conditions, and eliminates for instance effects due to conducting the experiment at different moments. In the ‘no anchor condition’ the participants received a meaningless letter CAPTCHA. The participants assigned to any of the other four conditions encountered at least one CAPTCHA with an anchor value.

The ‘standard anchoring effects condition’ linked to Q2 consisted of a CAPTCHA (as presented in Figure 4), a comparative judgment, and an absolute judgment.

Comparative judgment (related):

"Do you think the price (as set by the manufacturer and in euro) of the smartphone that you have been examining, is greater than or less than 532?"

Absolute judgment (1):

"What is your best guess of the price of the smartphone that you have been examining? (as set by the manufacturer and in euro)"

The 'unrelated comparison condition' is similar to the 'standard anchoring effects condition', except for the comparative judgment. To answer Q3, an unrelated factual question – acting as an unrelated comparative judgment – was used.

Comparative judgment (unrelated):

"Do you think the height of the Atomium of Brussels (in meter) is greater than or less than 532?"

In the calibration study the participants had to estimate the height of the Atomium to make sure that most participants know that the Atomium is smaller than 532 meters. If the unrelated factual question is easy enough it can also be used to grant consumers access to the website, as if this question also tries to tell computers and humans apart. It is however, crucial, that consumers do not know the exact height of the monument with certainty because this could act as strong anchor. If, for example, a comparative judgment with the year of the Battle of Hastings were used, many people would actively think of 1066 and would probably be influenced by this number instead of the chosen anchor value.

The fourth and fifth condition are the most important, because they create the basic anchoring effects to answer Q1. There are no comparative judgments in these conditions, but the CAPTCHA part is more elaborate. In the 'copy condition' the participants need to copy three regular CAPTCHAs with similar anchor values (543, 521, and 532). In the fifth condition or 'computation condition', the participants are three times asked to compare two CAPTCHA values and to copy the biggest of the two numbers (see Figure 5).

Repeating the anchors and demanding mental processing – copying or computation – both aim to maximize the chances of obtaining sizable basic anchoring effects. The existing literature

does not propose a standard on how many repetitions there should be to obtain basic anchoring effects. Wilson et al. (1996) find that seven times is too few, and thirty-five enough. Brewer and Chapman (2002) confirmed these numbers in their replication of Wilson et al. (1996). Mussweiler and Englich (2005) repeat their subliminal anchors ten times, and Wu et al. (2008) three times[49]. When designing a business application, the number of repetitions should be minimized to limit annoyance. Thirty-five is therefore not an option. Since the use of CAPTCHAs generates more attention because of the font disfluency (Mead & Hardesty, 2018) and given the inherent processing effort due to copying the number, I chose to work with three repetitions.

To answer Q4, the price of a more or less similar smartphone is shown after the first absolute judgment (see Figure 7). This reference price is immediately followed by a second absolute judgment, allowing the participants to change the estimate they provided in the first absolute judgment. The reference price and second absolute judgment were presented to the participants in all conditions. In other words, the analysis of Q4 will be based on a within-subject or repeated-measures design.

Absolute judgment (2):

"Now that you have found the price for a more or less similar smartphone,
what is your new best guess of the price of the first smartphone?
(as set by the manufacturer and in euro)"

5.5 Population and sampling method

The participants were 232 business or economics students at the *Vrije Universiteit Brussel* between 18 and 25 years old. I chose to work with students because most studies on anchoring use a student samples (Furnham & Boo, 2011)[50]. The age window was imposed to enhance the homogeneity of the sample. Excluding working students, for instance, limits disparities in available income. The participants were randomly allocated to the five experimental conditions

[49] These two studies should not be used as guidelines head over heels. Subliminal anchoring is very specific, and the study of Wu et al. (2008) may only have worked with three repetitions because of the other psychological phenomena in place (see section 4.3).

[50] Of the all the empirical studies on anchoring effects that are mentioned in the present dissertation, only Northcraft & Neale (1987) and Englich et al. (2006) performed experiments on experts. All others used graduate or MBA students.

(see 6.1). Descriptive statistics on gender, field of study, and age of the participants can be found in Appendix B.

The experiments took place during six different lectures[51] given to six distinct student groups between March 12th, 2018 and March 20th, 2018[52]. This made it easy to ensure that participants did not participate multiple times. At the same time, it reduced the possibility that participants would be aware of the purpose of the experiment and the CAPTCHA manipulations because they overheard fellow students[53]. Accidentally including these two student groups could not only contaminate the results because of their suspicious attitude towards the CAPTCHAs, but also because an earlier formulated or overheard price estimate could become a – supposedly – strong source of anchoring effects. Note that, for similar reasons, the 232 participants differ from the students who participated in the calibration study.

The anchoring effect has several moderating factors (see Furnham and Boo (2008) for an overview), but most only produce small differences and were therefore not used to preclude students from participating. Rather the experiment controls for the most important moderating factors, i.e. and product familiarity, knowledge about the target value (Van Exel et al. 2006; Wilson et al., 1996). Based on information that was retrieved from the participants at the end of the survey, we also examine the confidence participants have in their own estimates, how they processed the anchor value, and whether or not they perceived some manipulation from the CAPTCHA. Because these factors are intertangled, a segmentation of the population into subgroups with specific quotas on the number of participants was not possible. Therefore, a convenience sample was used, which is once again common in the existing literature and not problematic due to the absence of strong moderating factors and the random allocation of the participants to the experimental conditions. The number of participants per condition is discussed at the beginning of chapter 6, right before the results of the experiment are presented.

[51] This made participating in the experiment an implicit obligation. The participants could win a cinema ticket, but this incentive was not contingent upon the accuracy of the participant's estimates.
[52] On March 12th a pilot study with 17 participants was conducted. The pilot showed that there was no reason to change the survey or elements peripheral to the experiment, the results for these 17 participants were included in the final data set.
[53] At the beginning of each experiment, I asked students to indicate whether they had already participated and whether they were informed by other students about the experiment. After the experiment, I also checked the personal data of the students to rule out students who participated more than once – of which there were none.

6 Results

In this chapter, the four research questions will be answered based on the data that was collected by means of the experiment set out in Chapter 5. At the beginning of the experiment, every participant was randomly assigned to one of the experimental conditions. Initially, I aimed for an equal number of participants in each experimental condition, but due to unusable answers[54], the experimental design became unbalanced, with the number of participants per condition ranging from 43 to 48 (see Table 5). In addition, another 49 participants were filtered out because they claimed they had identified either the target smartphone (see Figure 6) or the 'more or less similar' reference phone (see Figure 7). Whether the 49 participants identified the phones correctly is of minor importance[55], because it is likely that they based their price estimate on the price of the smartphone they had in mind. Since it is difficult to control for these effects – they are individual and not specific to one of the experimental conditions – the analysis uses only data of 183 participants instead of 232[56]. As can been seen in Table 5, the number of participants per condition lies between 33 and 40.

Table 5

Number of participants per experimental condition

	participants	participants who identified at least one phone	**used answers**
no anchor	45	12	**33**
standard anchoring	48	11	**37**
unrelated comparison	48	8	**40**
copy	43	10	**33**
computation	48	8	**40**
total	**232**	49	**183**

[54] 247 participants submitted complete answers, but 15 of them were removed from any analysis because they were older than 25 years and/or completed the experiment in a way that cannot be described as cooperative or sincere (e.g. estimating a smartphone price equal to €1).

[55] In reality, most of these participants were wrong. Only two and five participants correctly identified the brand of the target and reference smartphone, respectively.

[56] Surprisingly, including the 49 participants barely alters the mean price estimates and the results of the statistical analyses. This can be explained by the fact that they identified very different smartphones, with retail prices ranging from €200 to €800. Consequently, almost half of them (21 to be precise) provided and estimate below €400, while the other half (28) provided estimates above €400. There is no systematic error, which could have been the case if the majority of the 49 participants thought of a very expensive or very cheap smartphone.

Section 6.1 focuses on the impact of basic anchoring effects on WTP in an e-commerce setting. In section 6.2 the effectiveness of basic anchoring is compared the alternative set-ups, and, finally, we examine the durability of standard and basic anchoring effects in case consumers encounter a relevant anchor in section 6.3.

6.1 A practical application involving basic anchoring effects

A between-subjects one-way ANOVA on the log price estimates reveals a significant effect of the experimental condition ($F(4, 178) = 4.10$, $p = .003$)[57]. The results of a pairwise comparisons *t*-test[58] provided insight in the statistical significance of the differences between the means of particular experimental conditions (see Table 6 and Table 8). Although these results are related to the first three research questions (Q1, Q2, and Q3), section 6.1 focuses only on Q1: Can basic anchoring effects influence WTP in an e-commerce setting? As a reminder, basic anchoring effects are the effects of irrelevant anchors in a set-up where the decision maker is not forced to consider the anchor value as a possible target value. The results of the 'standard anchoring condition' and the 'unrelated comparison condition' are discussed in section 6.2.

Did basic anchoring effects influence the price estimates in the conducted experiment? Figure 8 shows that the mean price estimates in the 'copy condition' (N=33, M=383, SD=145) and the 'computation condition' (N=40, M=491, SD=201) are higher than the mean estimate in the 'no anchor condition' (N=33, M=356, SD=146)[59]. However, the price estimates of the copy condition are not significantly higher than those of the unanchored group ($p > .45$; see Table 6). A significant anchoring effect can only be found if the participants have to compare two CAPTCHAs and copy the biggest of the two numbers ($p < .01$). This element of computation is necessary, because copying a CAPTCHA three times – the manipulation in the copy condition – does not produce a significant anchoring effect. These results confirm the prediction made in

[57] Most studies on the anchoring effect use a logarithmic transformation to increase the normality of the data and the homogeneity of the variances, which are conditions that must be fulfilled before ANOVA can be applied. The log price distributions are normally distributed (Shapiro-Wilk normality test on the residuals of the ANOVA model, $W = 0.99$, $p = .07$) and have a common variance (Levene's Test for homogeneity of variance, $F(4, 178) = 1.28$, $p = .28$). Every time ANOVA is used in this chapter, these assumptions are fulfilled.

[58] The Benjamini-Hochberg correction for multiple testing is used to reduce Type I errors. It is less conservative than the traditional Bonferroni-procedure and controls the false discovery rate (Benjamini & Hochberg, 1995). It is also the proposed correction in the R statistical environment (R Core Team, 2018).

[59] The analysis is focused on the means of the conditions, rather than the medians. Due to the small samples and the fact that many of the participants provided estimates that are multiples of 50 (numbers ending on '00' or '50'), the medians of each condition are very sensitive to adding or removing even a single estimate.

section 4.2.1: basic anchoring effects are able to produce significant increases in the price estimates, but this depends on the experimental paradigm.

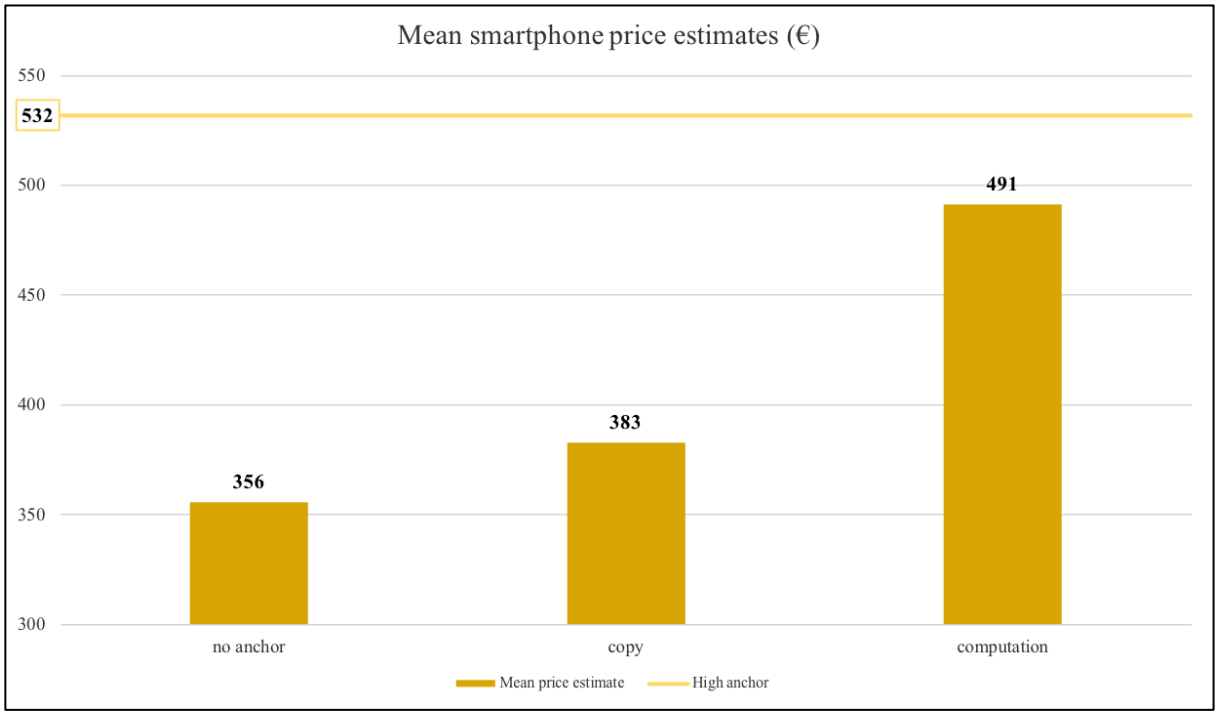


Figure 8. A comparison of the mean price estimates in the different experimental conditions.

Table 6

p-values of the pairwise comparisons t-test on the log estimates

	no anchor	copy
copy	.492	/
computation	.006 **	.024 *

Note. Significance codes: p <.001 '***', p <.01 '**', p <.05 '*', p <.10 '.'.

Note that the mean of the computation condition is not only significantly higher than the mean of the unanchored condition, but also than that of the copy condition, albeit this time at the 0.05 level. This shows clearly that the effect of the copy condition is smaller than that of the computation condition. Three differences between both conditions can account for this finding.

Presumably the required processing effort – which is higher in the computation condition (see section 2.5.3) – plays the biggest role. Copying three CAPTCHAs apparently does not

require a sufficiently big mental effort. Another difference between the two conditions lies in the number of anchors. In both conditions, the participants had to copy three CAPTCHAs, but in the computation condition each time *two* numeric CAPTCHAs were presented. Consequently, the computation condition contains three additional anchors, or six in total (see Table 4)[60,61].

The final difference between the copy and the computation condition is related to a drawback of using CAPTCHAs. As mentioned in section 4.3, when participants copy the anchor values, they can read the CAPTCHA as three separate numbers (five-three-two) rather than as one value (five hundred thirty-two). The comparison exercise in the computation condition forces the participant to read the CAPTCHAs as one value[62], but there is no similar mechanism in the copy condition[63]. To control for this effect, the participants were asked after the experiment to indicate whether they had read the CAPTCHA as separate numbers or as one number. More than half (60%) of the participants in the computation condition read the CAPTCHA as a number, compared to only one third (36%) in the copy condition (see Table 7). These figures would seem to support the hypothesis that reading separate numbers is a possible explanation of the absence of an anchoring effect in the copy condition. If this were true, however, one would also expect differences within each condition between the participants who read separate numbers and those who read one number. Descriptive and inferential statistics do not show these expected differences[64]. Although there is no conclusive evidence, it is possible that reading the CAPTCHA as separate numbers does not lead to lower price estimates. This need not be surprising, since the participants may unconsciously have processed the CAPTCHA as one number unconsciously after all, as is the case in the subliminal anchoring paradigm.

---

[60] Not three additional anchor *values*, because 532, 521, and 543 are used twice. The details can be found in Appendix A.

[61] It would be interesting to test this explanation by asking participants to copy three anchor values twice (i.e. a more elaborate copy condition) and comparing the effect with the demonstrated effect of the computation condition.

[62] This is no absolute guarantee. Participants might notice that both numbers start with '5' and compare only the last two digits, maybe even one by one (see Figure 3).

[63] In the standard anchoring condition and the unrelated comparison condition, the participants read 532 as one number at least once, namely in the comparative question.

[64] The estimates of the participants who did answer 'no idea' when asked how they had read the numeric CAPTCHA, are excluded from this analysis. A two-way ANOVA on the log price estimates revealed no significant main effect for the way the CAPTCHA was read ($F(1, 108) = 0.75$, $p = .39$) nor an interaction effect with the experimental condition ($F(3, 108) = 0.24$, $p = .87$). A Type-III sums of squares method was used because of the unbalanced design. Because of the conceptual differences between the conditions in the impact of how the CAPTCHA was read, I conducted Mann-Whitney-Wilcoxon rank sum tests. They showed no difference between the 'one number' means and 'separate numbers' means (p's $> .15$), except in the standard anchoring condition, where the mean of the participants who read the CAPTCHA as separate (!) numbers is higher ($U = 63$, $p = .09$).

Table 7

How participants have read the numeric CAPTCHA (percentage of participants per condition)

	one number *(532)*	separate numbers *(5-3-2)*	no idea
standard anchoring	**41**%	38%	21%
unrelated comparison	23%	**40**%	37%
copy	36%	**52**%	12%
computation	**60**%	23%	17%

Note. The largest group in each condition is indicated in bold.

As pointed out before, L (see Table 6). However, the stylized violin plots in Figure 9 do show differences between the distributions of the copy condition and the 'no anchor condition'[65]. A descriptive assessment of the distributions based on the concept of stochastic dominance provides a more nuanced understanding of the effects of the experimental conditions.

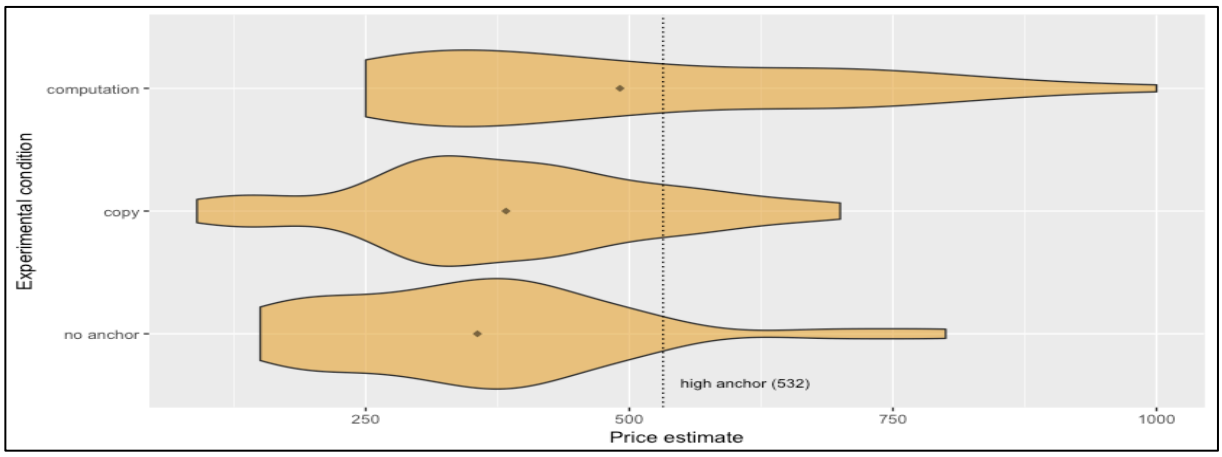


Figure 9. Stylized price estimate distributions for each experimental condition. *Note.* The little diamonds indicate the mean price estimate.

Koçaş and Dogerlioglu-Demir (2014, p.143) define stochastic dominance as follows: "For any two cumulative distribution functions (CDFs), $F_j(x)$ first-order stochastically dominates $F_i(x)$ iff $F_j(x) \leq F_i(x)$, for all x". Figure 10 shows the smoothed empirical cumulative distribution functions for the experimental conditions. Consistent with Koçaş and Dogerlioglu-Demir

[65] Although a couple of low estimates seem to account for the low mean in the copy condition, this is not the case. The low values are no outliers in the distribution of all price estimates, nor in the distribution of the copy condition. Even if they were outliers, this would not be due to insincere answers (as these have been filtered out) or errors of measurement.

(2014), a 10% truncation of the tails is used. Within these boundaries – indicated by the dotted horizontal lines – the pink CDF of the computation condition is positioned to the right of the red unanchored CDF. This implies that the percentage of participants who estimated a price lower than any given price is higher in the no anchor condition than in the computation condition. Hence, the computation condition stochastically dominates the 'no anchor' condition, which is a higher-order dominance than the difference between the means. Note that the main part of the CDF of the copy condition is also positioned to the right of the red unanchored CDF, even though there is no significant anchoring effect in the copy condition.

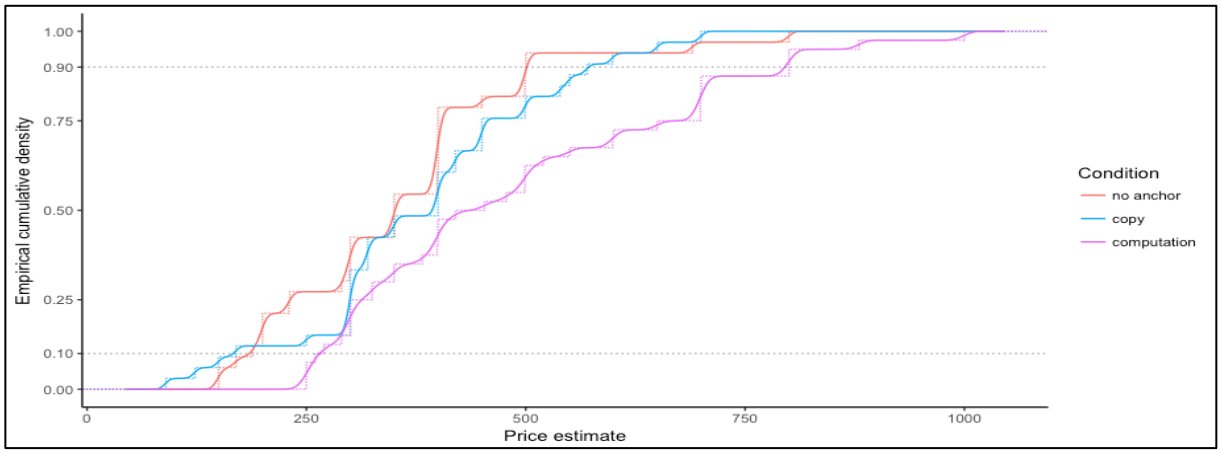


Figure 10. (Smoothed) Empirical cumulative distributions for the price estimates of each experimental condition. *Note.* The dotted horizontal lines represent the 10% truncation at both tails.

A short example indicates the importance of stochastic dominance when examining the influence on WTP and revenue (see section 3.2.2). The empirical CDFs in Figure 10 are assumed to represent the CDFs of the whole consumer population. For simplification purposes, I assume that the price estimate provided by a participant is equal to his IRP and his WTP[66]. If a seller wants to serve 25% of the consumers, he can charge a price of €400 (the 75th percentile of the unanchored distribution) for the smartphone. Demanding consumers three times to compare two CAPTCHAs and to copy the largest allows the seller to charge €662 while serving the same number of consumers. Simply copying three CAPTCHAs enables pricing at €450, which is still a substantial revenue increase. As stated in section 3.2.2, changing the IRP of consumers can also raise revenue at a given price, i.e. by selling more smartphones. Suppose

[66] Although the price estimate could be equal to the IRP, it is rather unlikely that it is also equal to the WTP. Assuming that a consumer's WTP is characterized (ceteris paribus) as a strictly increasing function of his IRP is more realistic, but does not change the essence of the example.

that the seller has decided upon a price of €499.99. If he uses the computation procedure, almost four out of ten (38%) consumers will buy. This is much more than without anchoring (12%) or with a copy procedure (21%). Using a stochastically dominant procedure implies that the demonstrated revenue increases do not only materialize for a target of 75% or a predetermined selling price €499.99, but for every percentage and every price.

In conclusion, basic anchoring effects can effectively influence the price estimates of consumers. Copying a CAPTCHA does not produce a significant anchoring effect, but asking consumers to compare numeric CAPTCHAs and to copy the biggest of the two numbers clearly does. The distribution of the price estimates showed that if a seller asks his consumers to perform this 'computation' – under the guise of trying to tell humans and computers apart – his revenue will increase substantially.

6.2 Alternative set-ups with irrelevant anchors

Section 6.1 demonstrated that basic anchoring effects can influence WTP in an e-commerce setting. However, alternative set-ups with irrelevant anchors – in this case the standard anchoring condition and the unrelated comparison condition – might be more effective (cf. Q2 and Q3). If this is true, it would pose a trade-off between a low-annoyance, low-suspicion, but small basic anchoring effect and a more annoying and suspicious, but also more impactful comparative procedure. The results of the experiment show that no such trade-off exists: basic anchoring effects can be (at least) as effective as the alternative set-ups.

Figure 11 shows that the mean price estimates in the standard anchoring condition (N=37, M=447, SD=146) and the unrelated comparison condition (N=40, M=417, SD=142) lie between those of the copy and the computation condition. As expected, the robust standard anchoring paradigm produces a significant anchoring effect (p <.05; see Table 8). The mean of the standard anchoring condition is also higher than that of the copy condition (be it only marginally significant, p <.10)[67]. However, the computation condition demonstrates that basic anchoring effects are definitely not smaller than standard anchoring effects (p >.45). This is also supported by the stylized distributions shown in Figure 12.

[67] To examine differences between the four anchoring conditions (N=150), I performed an additional one-way ANOVA on the log price estimates of those conditions (F(3, 146) = 2.88, p =.004). The results of the subsequent pairwise comparisons t-test (Benjamini-Hochberg correction) are similar to those presented in Table 8, except for the difference between the means of the standard anchoring condition and the copy condition, which is no longer significant (p =.130).

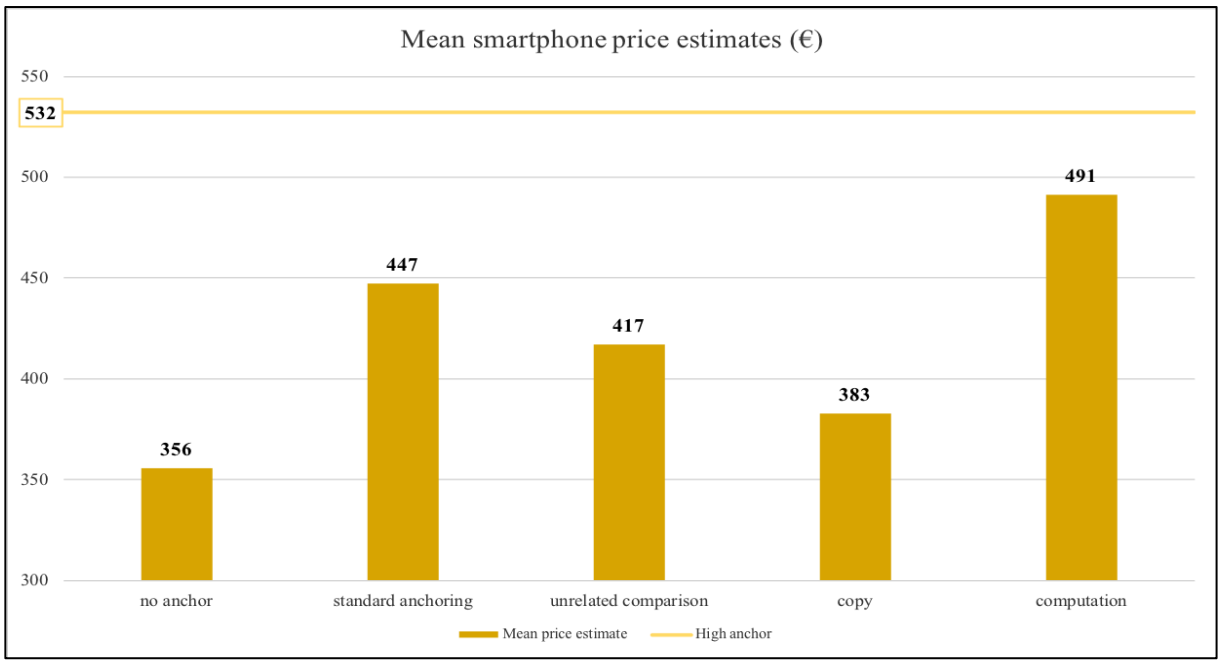


Figure 11. A comparison of the mean price estimates in the different experimental conditions.

Table 8

p-values of the pairwise comparisons t-test on the log estimates

	no anchor	standard anchoring	unrelated comparison	copy
standard anchoring	.024 *	/	/	/
unrelated comparison	.095 .	.492	/	/
copy	.492	.095 .	.292	/
computation	.006 **	.492	.190	.024 *

Note. Significance codes: p <.001 '***', p <.01 '**', p <.05 '*', p <.10 '.'.

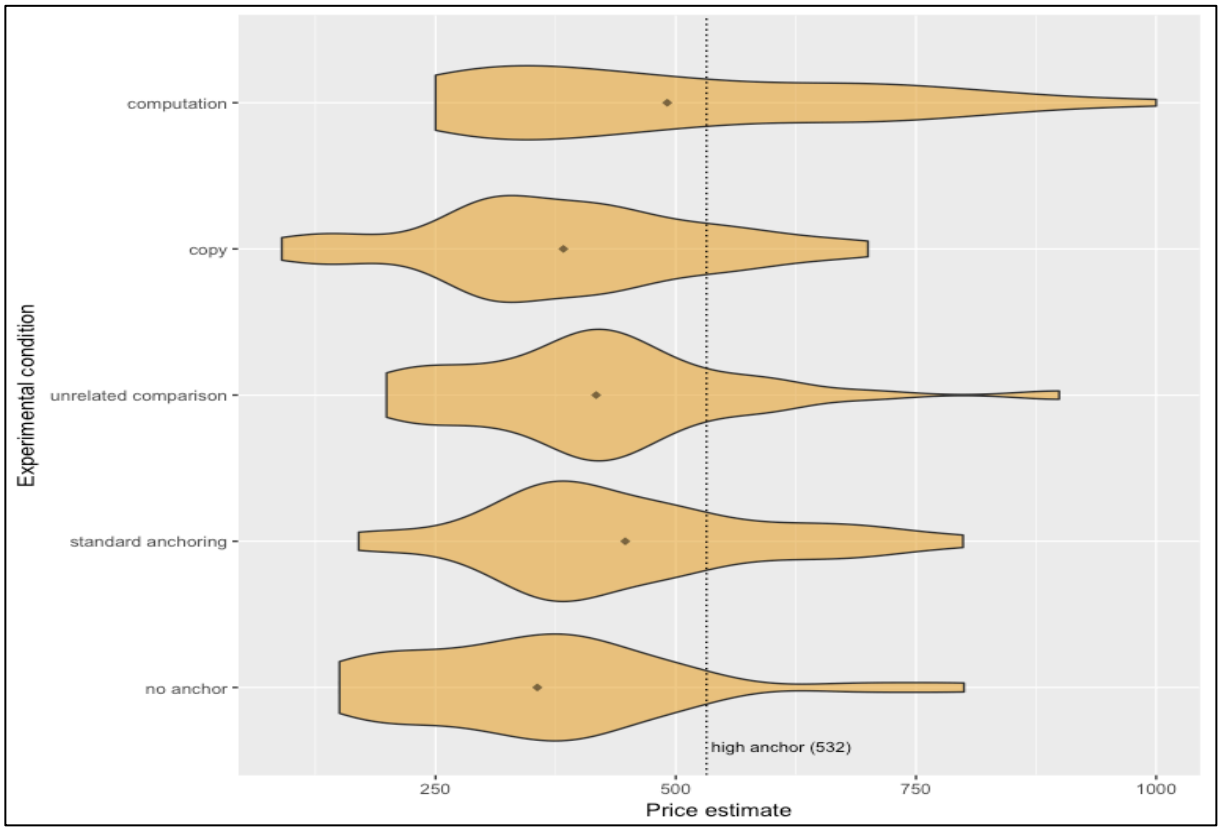


Figure 12. Stylized price estimate distributions for each experimental condition. *Note.* The little diamonds indicate the mean price estimate.

Using an unrelated comparative question also produces an anchoring effect, but this is only weakly significant (p <.10)[68]. The effect of the unrelated comparison is not significantly smaller than those of the standard anchoring (p >.45) and computation (p >.15) condition, but it is not significantly larger than the effect of the copy condition either (p >.25). It seems that using standard and basic anchoring (with a computation element) is to be preferred over the use of an unrelated comparative question.

Note that all the significant differences between the mean price estimates of the experimental conditions (see Table 8) translate into stochastic dominance of one condition over another (see Figure 13). As demonstrated in section 6.1, the implications of stochastic dominance are far from trivial in the context of influencing WTP.

[68] The effect is not significant (p =.211) if the 49 participants who identified one of the smartphones are included. This is the only important implication of excluding the 49 participants.

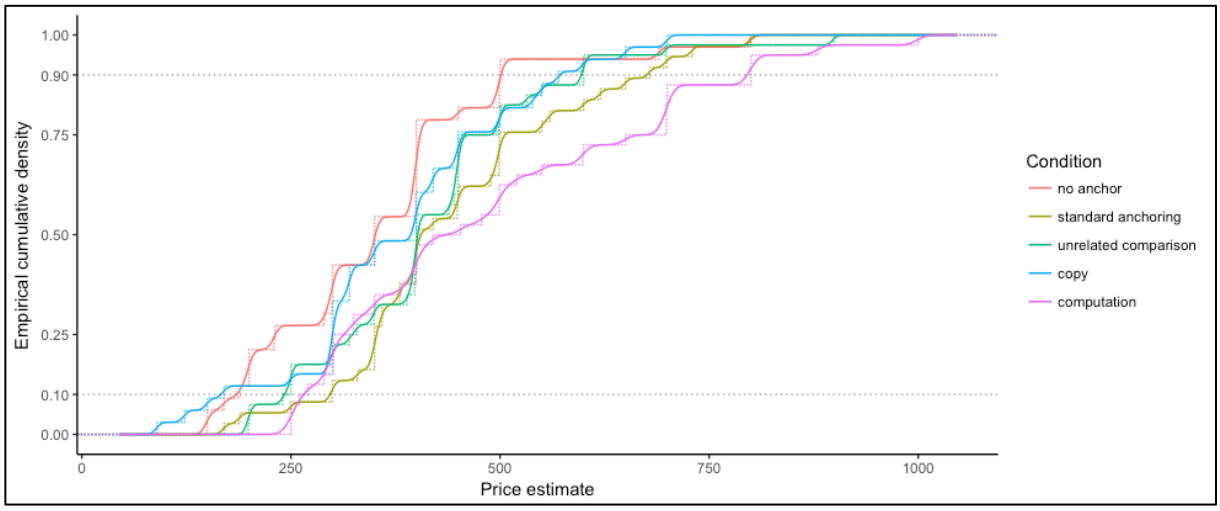


Figure 13. (Smoothed) Empirical cumulative distributions for the price estimates of each experimental condition. *Note.* The dotted horizontal lines represent the 10% truncation at both tails.

The effectiveness of an anchoring procedure is obviously important, but in practice sellers will not be eager to use any technique that causes suspicion, even if the effectiveness is not compromised by this suspicion. To determine how much suspicion each condition caused, I asked the participants whether they thought that the CAPTCHA had influenced their price estimates. One out of ten of the participants in both the copy (9%) and computation (10%) condition answered affirmatively, which is noticeably higher than in the standard anchoring (3%) and unrelated comparison (5%) condition. However, this does not mean that the inclusion of a comparative judgment causes less suspicion, because for these participants it is the comparative judgments that seems suspicious, rather than the CAPTCHA. Surprisingly, almost one out of five (18%) of the participants in the no anchor condition believed copying a meaningless letter CAPTCHA ('*eyr*') had influenced their judgments. This implies that the suspicion levels are not so high in the computation and copy condition after all. Moreover, it illustrates that the percentages presented in this paragraph are probably higher than will be the case in a real setting, because participants become more suspicious when they are explicitly asked whether they think that a CAPTCHA had influenced them. In short, the computation procedure does not seem exceptionally suspicious, which supports using the computation procedure rather than including a comparative judgment.

At the end of the experiment the participants were also asked whether or not they had heard about the anchoring phenomenon. Surprisingly, only 3 of the 21 students that had some

understanding of anchoring effects, answered that the CAPTCHA had influenced their judgments. This may indicate that understanding the anchoring phenomenon is not enough to suspect a CAPTCHA. Additionally, suspicion does not seem enough to withstand the anchoring effects. A two-way ANOVA[69] on the log price estimates revealed no significant main effect of being suspicious ($F(1, 173) = 0.07$, $p = .78$), nor an interaction effect with the experimental condition ($F(1, 173) = 0.32$, $p = .87$). The main effect related to the experimental condition remained significant ($F(4, 173) = 3.89$, $p = .005$).

Knowledge is an important moderating factor of the anchoring effect. At the end of the experiment the participants were asked to indicate their knowledge of smartphone prices on a 9-point Likert scale (1: very poor; 9: excellent). The lengthy analysis – presented in Appendix C – demonstrated that knowledge has an impact on the price estimates of participants, but there is no evidence that the results in Table 8 can be explained by differences in knowledge between the participants of the different experimental conditions, nor by a different effect of knowledge per experimental condition.

After the participants provided their estimates, they also indicated their confidence in their price estimate on a 9-point Likert scale (1: not at all confident; 9: extremely confident). A Kruskal-Wallis test shows no significance difference between the means of these confidence scores due to the experimental condition ($\chi^2(4, N=183) = 0.38$, $p = .98$). Nor comparative judgments, nor copying three CAPTCHAs did lower their confidence in their own estimates.

In summary, basic anchoring effects can be at least as effective as standard anchoring effects. Although it seems that an unrelated comparative judgment can produce anchoring effects, these effects are clearly smaller. Finally, from a business perspective, it is definitely more practicable to ask consumers to compare CAPTCHAs than to include comparative judgment, which is required in the standard anchoring paradigm.

6.3 Relevance and durability of anchoring effects

Basic and standard anchoring effects can effectively influence price estimations, but Q4 addresses a justified concern: do these anchoring effects vanish when the participants are confronted with a suggested reference price?

[69] A Type-III sums of squares method was used because of the unbalanced design.

Introducing a reference price for a more or less similar smartphone (see Figure 7) had a dramatic impact, as is expected in the case of suggested-reference-point anchoring effects. It caused the vast majority of the participants (83%) to adjust their price estimates, almost exclusively towards the reference price[70]. The power of the reference price lies in the fact that most of the participants did perceive the reference price as trustworthy. This is demonstrated by the fact that the participants had significantly more confidence in the adjusted estimate than in their initial estimate[71]. The other participants (17%) did not act upon the new information because their initial estimates were already reconcilable with the reference price[72]. In the remainder of this section, 'second estimate' refers to the price estimate provided after seeing the reference price.

The impact of the reference price can also be seen in Figure 14, which compares the distributions of the initial price estimates with those of the second estimates. As expected, the estimates that were provided after the reference price are more concentrated around the mean. This supports the hypothesis that participants – on average – adjust away from their initial estimate towards the reference price (see section 4.2.4). Figure 15 shows how most of the differences between the CDFs of the experimental conditions disappeared. Only the computation condition and the copy condition stochastically dominate the no anchor condition. However, in both cases the differences are small compared to those in Figure 13.

[70] 138 of the 152 participants who adjusted their estimate, did this in the direction of the reference price. In the experiment, I asked the participants to explain why and how they changed their estimates (see Appendix A for an overview of the questions). Therefore, I know that the other 14 participants adjusted away because they questioned that the reference smartphone was 'more or less similar to the first smartphone'. Most of them believed that the dual camera – a feature specific to the first smartphone –justified a higher price than €367. They changed for instance their initial estimate from €400 to €450.

[71] I used a three-step statistical approach because there is no well-regarded non-parametric alternative for a two-way ANOVA in case of an unbalanced design. A paired Wilcoxon signed-rank test shows that the mean of the confidence in the second estimate is significantly higher than the confidence in the first estimate (Z = 3258, p <.001). This is also true in each experimental condition (p's <.05), except for the copy condition (Z = 169.5, p =.17). However, there does not seem to be an interaction between the increase in confidence and the experimental condition since a Kruskal-Wallis test showed no significant differences between the experimental conditions for the confidence in the second estimate ($\chi^2(4, N=183) = 6.47$, p =.17).

[72] The initial estimates of the 31 participants range from 300 to 600, but 22 of them are between 350 and 450.

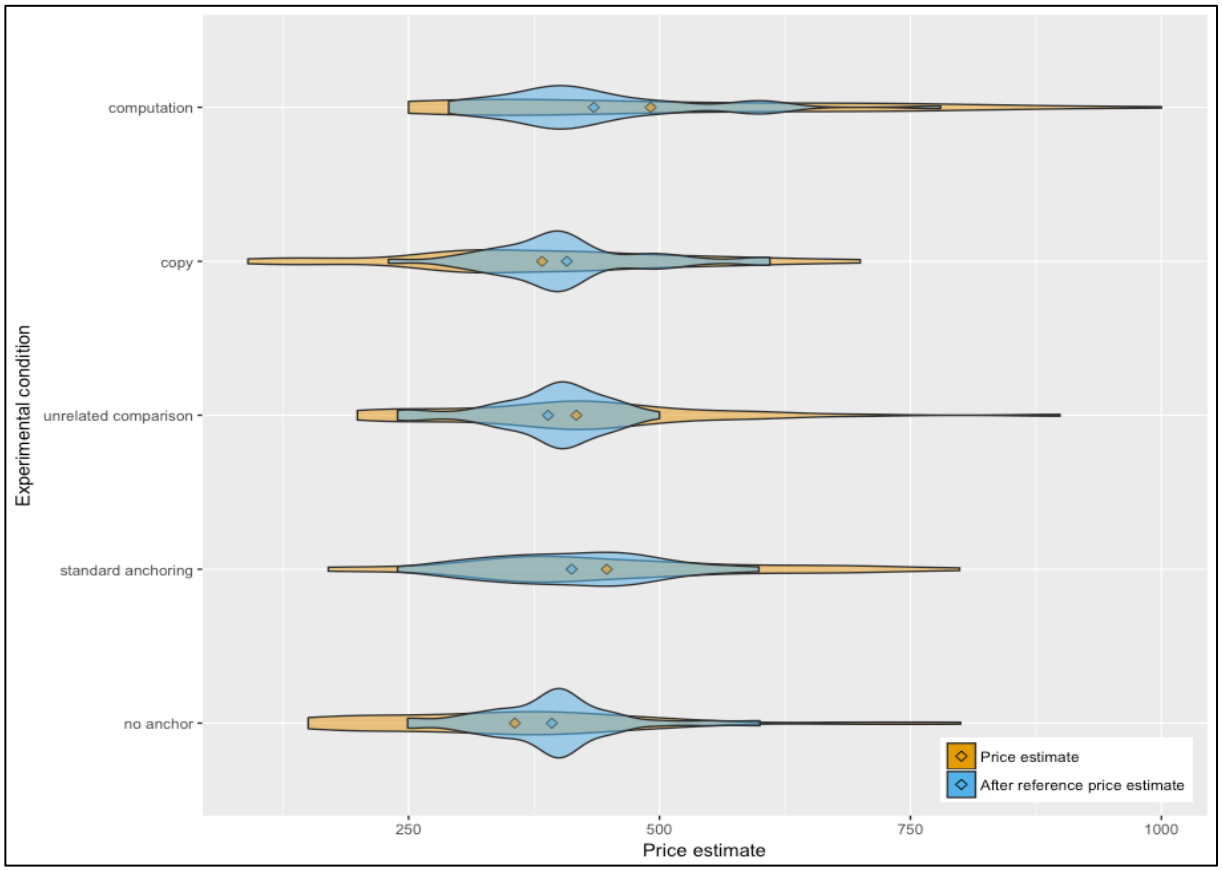


Figure 14. Stylized distributions of the price estimates before and after presenting the reference price. *Note.* The little diamonds indicate the mean price estimates.

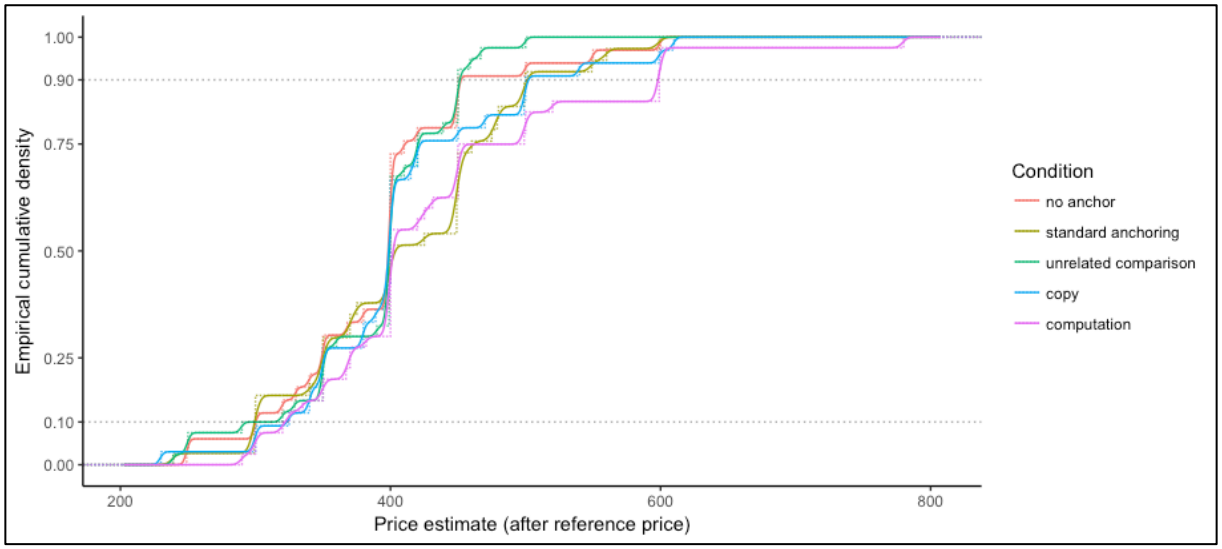


Figure 15. The (smoothed) empirical cumulative distributions for the price estimates of each experimental condition after presenting the reference price. *Note.* The dotted horizontal lines represent the 10% truncation at both tails.

The suggested-reference-point anchoring effect is demonstrated in Figure 16, which shows the differences between the means of the first and second estimates for all experimental conditions. The means of the three experimental conditions that produced significant anchoring effects (see sections 6.1 and 6.2) have become smaller. However, the adjustment towards the reference price is insufficient, since the second estimate's mean of the computation condition (N=40, M=435, SD=103) is still higher than those of the standard anchoring condition (N=37, M=412, SD=82) and the unrelated comparison condition (N=40, M=389, SD=59). Conversely, the means of the no anchor condition (N=33, M=393, SD=73) and the copy condition (N=33, M=408, SD=81) have increased due to the reference price[73,74].

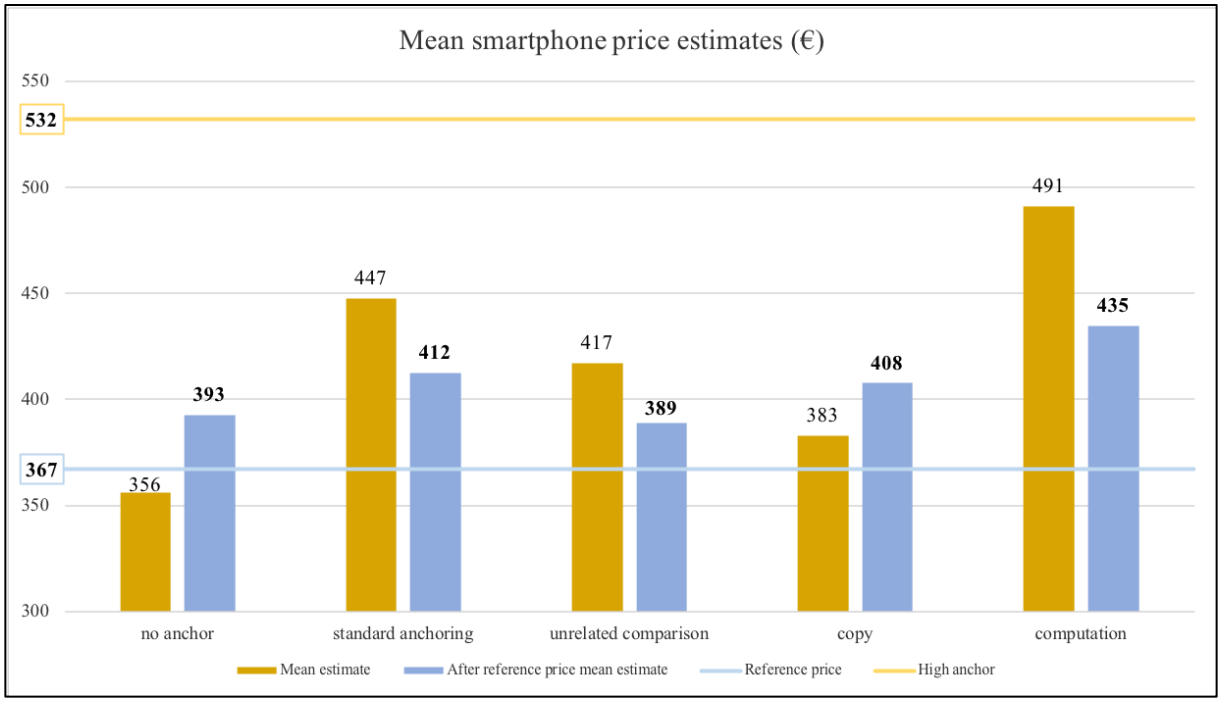


Figure 16. A comparison of the mean price estimates before and after presenting the reference price.

[73] Paired Wilcoxon signed-rank tests confirm that the means of the initial and second estimates are significantly different in the no anchor condition ($Z = 115.5$, $p = .05$), the standard anchoring condition ($Z = 406$, $p = .06$), and the computation condition ($Z = 372$, $p = .04$). There is no significant suggested-reference-point anchoring effect in the unrelated comparison condition and the copy condition ($p > .25$), presumably because the initial means of these conditions (417 and 383, respectively) are not so different from the reference price (367).

[74] At first sight, it seems odd that the mean of the no anchor condition has increased to a level above the reference price. Even more intriguing is the observation of a higher mean (after reference price) in the copy condition, for which the initial mean was already higher than the reference price. Both effects are related to the way in which the reference was introduced (see Figure 4). The reference smartphone has "more or less similar specifications (it has for instance no dual camera)". The participants evaluated this statement and decided that the dual camera made the first smartphone somewhat more expensive. This explains a focal point at €400 rather than exactly at the reference price of €367.

Although descriptive statistics indicate durable effects of the irrelevant anchor in the computation and standard anchoring condition, a Kruskal-Wallis test revealed no effect of the experimental condition on the second price estimates ($\chi^2(4, N=183) = 4.57$, $p = .33$)[75]. In other words, the initial standard and basic anchoring effects (see Table 8) have been wiped out by the suggested-reference-point anchoring effects.

This result seems to weaken the potential for a practical application involving basic anchoring effects, because consumers – and especially Internet shoppers – are often able to find reference prices. The implications of all findings presented in this chapter are discussed in Chapter 7, together with suggestions for future research.

[75] A Kruskal-Wallis rank sum test was conducted because log transforming the price estimates did not ensure normality (Shapiro-Wilk test, $W = 0.98$, $p = .03$).

7 General discussion

This chapter summarizes the main theoretical and managerial implications of the results that were presented in chapter 6. At the end of the chapter, I will dwell on the limitations of the present paper and the possibilities for future research.

7.1 Theoretical implications

In the present paper, I explored the influence of irrelevant anchors on price judgments and confirmed the existence of the anchoring effect: although the participants should have ignored the CAPTCHAs, these numeric values influenced the subsequent estimation of the price of a smartphone.

The robustness of the standard anchoring paradigm has been re-established, as well as the fragility of basic anchoring effects. A first basic anchoring procedure – copying three similar anchor values – did not produce a significant anchoring effect. But a procedure that requires more mental effort – comparing two numeric values and copying the biggest – did lead to sizable anchoring effects. Quite remarkably, the effects of this computation procedure are at least as big as standard anchoring effects. This demonstrates that basic anchoring effects can be (at least) as effective as standard anchoring effects, even though the existing anchoring literature deems standard anchoring to be more effective (see section 4.2.2).

Including an unrelated comparative judgment yielded rather small anchoring effects, which supports the hypothesis that these effects belongs to the family of fragile basic anchoring effects (see Table 3). The fact that an *unrelated* comparative judgment (in the unrelated comparison condition) produces a smaller anchoring effect than a *related* comparative judgment (in the standard anchoring condition) provides another valuable theoretical insight. In contrast to what Chapman & Johnson (2002) argue (see section 2.5.3) the robustness of the standard anchoring paradigm is not explained by the ability of a comparative judgment to generate sufficient attention and mental processing of the anchor value, but by the fact that it forces the decision maker to consider the anchor value as a possible target value. In the anchoring literature, the standard anchoring paradigm is sometimes inaccurately defined as a set-up with a comparative judgment. The presented result stresses that the inclusion of a comparative judgment is in itself not sufficient to obtain robust and sizable anchoring effects. What matters is asking the decision

maker to consider the anchor value as a possible target value[76]. These findings support the classification of anchoring paradigms that I proposed in Table 3 (see section 2.5.3).

The experiment also demonstrates the expected robustness of suggested-reference-point anchoring effects. Presenting a relevant anchor – in this case the reference price for a more or less similar smartphone – produced an effect strong enough to wipe out the initial standard and basic anchoring effects. Even though descriptive statistics support the hypothesis that the adjustment from the first estimate towards the reference price will be insufficient (see section 4.2.4), we have to conclude that relevance matters: reference prices with well-chosen anchor values pose a threat to the durability of standard and basic anchoring effects.

7.2 Managerial implications

At the beginning of chapter 2, two reasons were mentioned as to why studying cognitive biases is interesting. On the one hand, it can enable decision makers to 'debias' their own judgments. On the other hand, someone who understands these systematic – and therefore predictable – biases can exploit the flawed decision making of others.

As explained in section 2.5.4, debiasing is particularly difficult in the case of anchoring effects. The present paper shows, indeed, that a basic understanding of the anchoring phenomenon does not imply that a numeric CAPTCHA – not even three CAPTCHAs – causes suspicion. In addition, it has also been demonstrated that suspicion is no sufficient condition for a less affected judgment, let alone a debiased judgment (see section 6.2)[77].

This dissertation bears better news for everyone who aims to utilize the flaws in the decisions of others, and especially for sellers who want to take advantage of consumers' susceptibility to anchoring effects. The results show that anchoring effects can influence the internal reference price of consumers, which is an important determinant of their willingness to pay. Moreover, the rather theoretical idea of influencing the IRP with basic anchoring effects can be applied in practice for three reasons. First of all, because the basic anchoring effect turned out to be sufficiently effective. Therefore, it does not make sense to use the annoying and suspicious

[76] At least for an anchoring paradigm based on irrelevant anchors, since suggested-reference-point anchoring effects – for which there is no comparative judgment (see Table 3) – are a robust phenomenon.
[77] Still, one result is promising: an experimenter-provided relevant anchor can wipe out standard and basic anchoring effects. This could be helpful in a situation similar to the situation presented in Figure 3.

standard anchoring paradigm in a real e-commerce setting. Secondly, because the anchor values can be presented in a subtle way, namely in CAPTCHAs. And last but not least, because the seller can actually force his consumers to perform the required 'computation' under the guise of trying to tell computers and humans apart. In short, to increase his revenue, the only thing a seller needs to do, is asking his consumers three times to compare two CAPTCHAs and to copy the biggest of the two values. By means of the concept of stochastic dominance, section 6.1 aptly illustrated how sellers can potentially increase their revenue for every chosen selling price and every chosen quantity supplied.

The experiment illustrated one important problem of this procedure. Encountering a reference prices before the final purchase decision can wipe out the basic anchoring effect, even if the IRP was already imprinted. One solution for sellers, is to make it difficult for consumers to obtain a reference price. It is not possible to keep consumers from finding other prices, but the seller can make them look irrelevant, by influencing the individual's perception of context and value alignment (see section 2.3.1). If he succeeds, two basic anchoring effects will compete, instead of the intended basic anchoring effects and a powerful suggested-reference-point anchoring effect (see section 2.5.2).

7.3 Limitations and future research

Some of the limitations of the presented research are obvious. The sample is very specific, only one target product is examined, and although everything is done to ensure external validity, this study fails to create a real e-commerce environment and a real purchase experience. Future research should therefore examine the effects on other populations, for different target products, and hopefully even in a real e-commerce setting. In addition, future research is also required to replicate the findings of the present paper with other anchor values and CAPTCHA procedures.

This dissertation shares two other limitations with most studies on anchoring effects on WTP. First of all, I have to acknowledge that measuring the price estimates is only a proxy for the WTP. It is not entirely clear how the price estimate relates to the IRP, nor how a consumer's IRP translates into his WTP. In addition, it is possible that the IRP and WTP are only constructed or imprinted because the consumers had to answer a very explicit target question. In reality, a consumer will ask himself the question 'Should I buy at this particular price or not?'. In this case, the real price will probably have a much bigger influence on the outcome of the purchase decision than the numeric CAPTCHAs.

Although there are probably many more limitations related to the business application, I would like to highlight two important ones. It is not possible for the seller to know what a consumer will want to buy when he visits the website. This would require that the CAPTCHAs contain optimal anchor values for each product that is sold on the website, which is hardly feasible in reality. Finally, the future is not so bright for the CAPTCHAs that were used in this dissertation. Today, these traditional CAPTCHAs are being replaced by CAPTCHAs that demand the consumer to indicate all the images that contain, for example, mountains. Although this might be an opportunity to prime the consumer, it implies that hiding numeric values in CAPTCHAs is no longer possible.

8 Conclusion

The present paper contributes to the anchoring literature in several ways. Firstly, because it sets out two original conceptual frameworks. The first is a categorization of the different anchor types based on the newly introduced concepts of context and value alignment. The proposed classification of the major anchoring paradigms, for its part, has also been validated. To obtain robust anchoring effects, the relation between the comparative judgment and the target value matters.

Moreover, the study is of value because it is one of the few experiments that compare standard and basic anchoring effects, that confront irrelevant anchors with relevant anchors, that look at distributions rather than measures of central tendency, etc. Although it seems self-evident, combining the influence of anchors on willingness to pay with Thaler's (1985) transaction utility is unprecedented. As far as I know, this dissertation is the only paper – other than Wu et al. (2008) – that examines and finds a practical application for basic anchoring effects. An application that is based on an innovative, effective, and promising way to present anchors in an e-commerce setting – numeric CAPTCHAs – and demonstrates that basic anchoring effects can be used to influence the price judgments of consumers. Let me state this again: *basic anchoring effects can be used to influence the price judgments of consumers*. At the end of this paper, one might have forgotten how astonishing and frightening it is that a simple CAPTCHA can be used to influence purchase behavior.

Standard economic theory ignores many powerful psychological phenomena. However, in our daily lives, in a world of Humans, most consumers and sellers do exactly the same. They consider a CAPTCHA to be irrelevant, and apparently they are wrong. Although I do not aim to arm sellers with this knowledge more than I want to inform customers, I secretly hope – from an academic point of view – that this paper encourages a seller to experiment with basic anchoring effects in a real e-commerce setting. And if he should succeed and complacently thinks about the unawareness of his consumers and competitors, I hope to hear him mumble:

Gotcha with a CAPTCHA.

References

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent Arbitrariness": Stable Demand Curves without Stable Preferences. *The Quarterly Journal of Economics*, *118*(1), 73-105.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). An experimental study of some stochastic models for wagers. *Behavioral Science*, *8*(3), 199-202.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300.

Brewer, N. T., & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making*, *15*(1), 65-77.

Cervone, D., & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology*, *50*(3), 492-501.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the Irrelevant: Anchors in Judgments of Belief and Value. In T. Gilovich, D.W. Griffin, D. Kahneman (Eds.), *The psychology of intuitive judgment: Heuristics and biases* (pp. 120-138). New York: Cambridge University Press.

Cialdini, R. B. (2007). *Influence: The Psychology of Persuasion* (Rev. ed.). New York: Collins.

Costa, D. F., de Melo Carvalho, F., de Melo Moreira, B. C., & do Prado, J. W. (2017). Bibliometric analysis on the association between behavioral finance and decision making with cognitive biases such as overconfidence, anchoring effect and confirmation bias. *Scientometrics*, *111*(3), 1775-1799.

Critcher, C. R., & Gilovich, T. (2008). Incidental Environmental Anchors. *Journal of Behavioral Decision Making*, *21*(3), 241-251.

Davis, H. L., Hoch, S. J., & Ragsdale, E. E. (1986). An anchoring and adjustment model of spousal predictions. *Journal of Consumer Research*, *13*(1), 25-37.

Englich, B., & Mussweiler, T. (2001). Sentencing Under Uncertainty: Anchoring Effects in the Courtroom. *Journal of Applied Social Psychology*, *31*(7), 1535-1551.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin*, *32*(2), 188-200.

Epley, N., & Gilovich, T. (2010). Anchoring unbound. *Journal of Consumer Psychology*, *20*(1), 20-24.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288-299.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35-42.

Glöckner, A., & Englich, B. (2015). When Relevance Matters: Anchoring effects can be larger for relevant than for irrelevant anchors. *Social Psychology*, *46*(1), 4-12.

Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, *20*(2), 85-116.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, *9*(1), 90-98.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of Anchoring in Estimation Tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161-1166.

Jones, B. D. (2002). Bounded rationality and public policy: Herbert A. Simon and the decisional foundation of collective choice. *Policy Sciences*, *35*(3), 269-284.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Koçaş, C., & Dogerlioglu-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.

Mandel, N., & Johnson, E. J. (2002). When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices. *Journal of Consumer Research*, *29*(2), 235-245.

Mead, J. A., & Hardesty, D. M. (2018). Price Font Disfluency: Anchoring Effects on Future Price Expectations. *Journal of Retailing*, *94*(1), 102-112.

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, *122*(1), 69-79.

Monroe, K. B. (2003). *Pricing: Making Profitable Decisions* (3rd ed.). New York: McGraw-Hill.

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes*, *98*(2), 133-143.

Mussweiler, T., & Strack, F. (2001). Considering the Impossible: Explaining the Effects of Implausible Anchors. *Social Cognition*, *19*(2), 145-160.

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*(1), 84-97.

Peterson, R. A., & Merino, M. C. (2003). Consumer Information Search Behavior and the Internet. *Psychology & Marketing*, *20*(2), 99-121.

R Core Team (2018). R: *A Language and Environment for Statistical Computing* [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business Press.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99-118.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review*, *69*(4), 493-513.

Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, *31*(3), 681-690.

Simonson, I., & Tversky, A. (1992). Choice in Context: Tradeoff Contrast and Extremeness Aversion. *Journal of Marketing Research*, *29*(3), 281-295.

Slovic, P. (1995). The Construction of Preferences. *American Psychologist*, *50*(5), 364-371.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment. *Organizational Behavior and Human Performance*, *6*(6), 649-744.

Strack, F., & Mussweiler, T. (1997). Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility. *Journal of Personality and Social Psychology*, *73*(3), 437-446.

Switzer, F. S., & Sniezek, J. A. (1991). Judgment Processes in Motivation: Anchoring and Adjustment Effects on Judgment and Behavior. *Organizational Behavior and Human Decision Processes*, *49*(2), 208-229.

Thaler, R. H. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, *4*(3), 199-214.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. New York: W.W. Norton & Company.

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, *112*(1), S164-S187.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453-458.

Van Exel, N. J. A., Brouwer, W. B. F., van den Berg, B., & Koopmanschap, M. A. (2006). With a little help from an anchor: discussion and evidence of anchoring effects in contingent valuation. *The Journal of Socio-Economics*, *35*(5), 836-853.

Varian, H. R. (2014). *Intermediate Microeconomics: A Modern Approach: Ninth International Student Edition*. New York: W.W. Norton & Company.

Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, *16*(1), 8-37.

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

Wilson, T. D., & Brekke, N. (1994). Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations. *Psychological Bulletin*, *116*(1), 117-142.

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*(4), 387-402.

Wong, K. F. E., & Kwong, J. Y. Y. (2000). Is 7300 m Equal to 7.3 km? Same Semantics but Different Anchoring Effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 314-333.

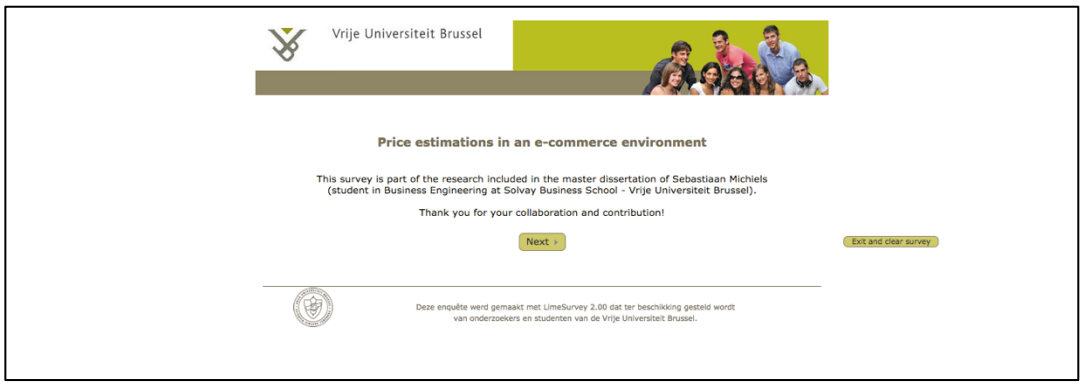
Wu, C. S., Cheng, F. F., & Lin, H. H. (2008). Exploring anchoring effect and the moderating role of repeated anchor in electronic commerce. *Behaviour & Information Technology*, *27*(1), 31-42.

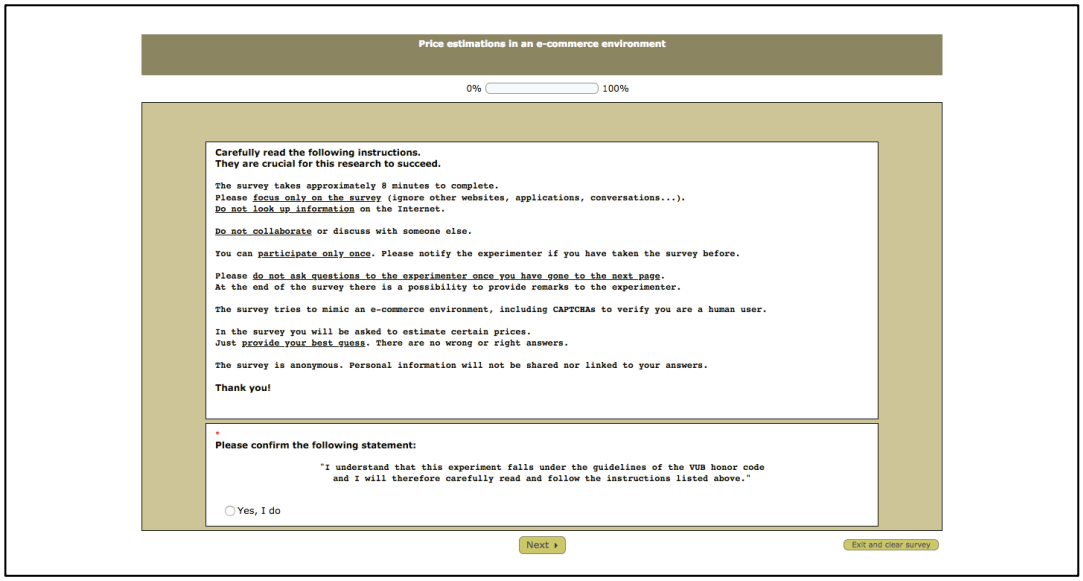
Wu, C. S., Cheng, F. F., & Yen, D. C. (2012). The role of Internet buyer's product familiarity and confidence in anchoring effect. *Behaviour & Information Technology*, *31*(9), 829-838.

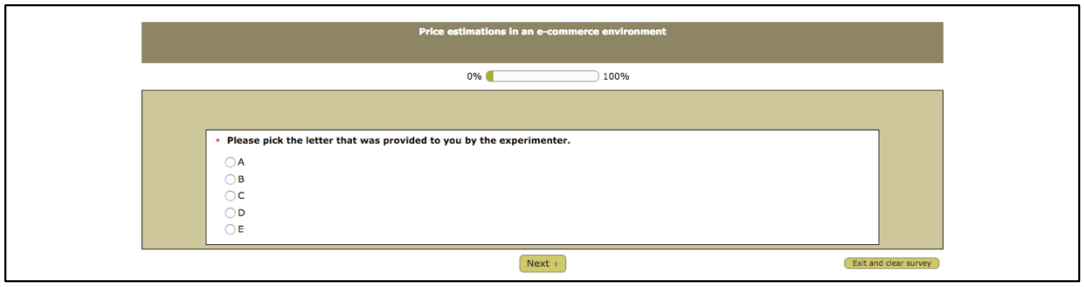
Appendix

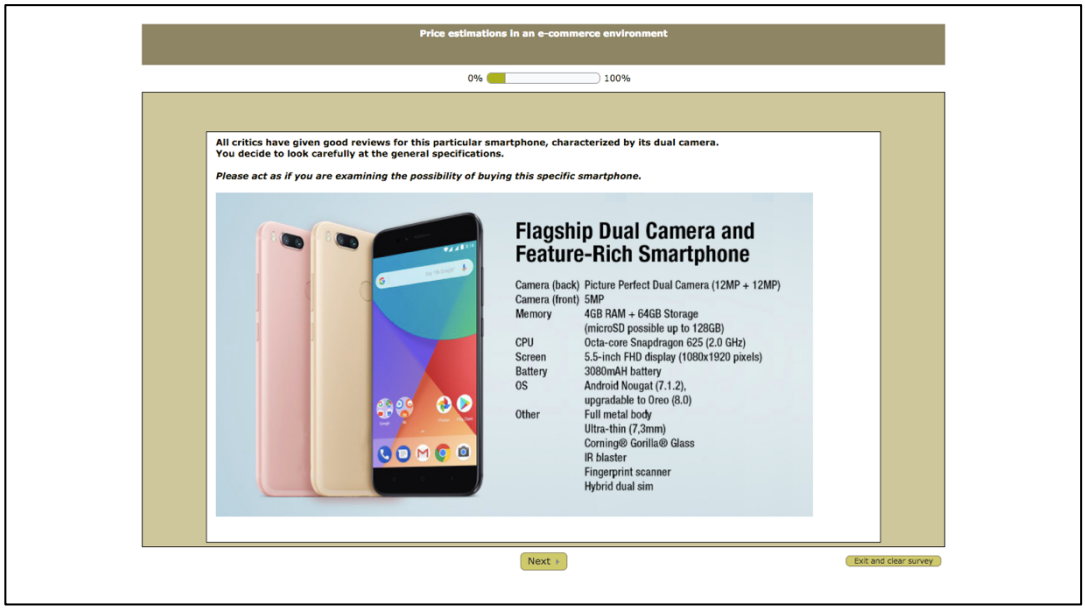
A Survey

All conditions

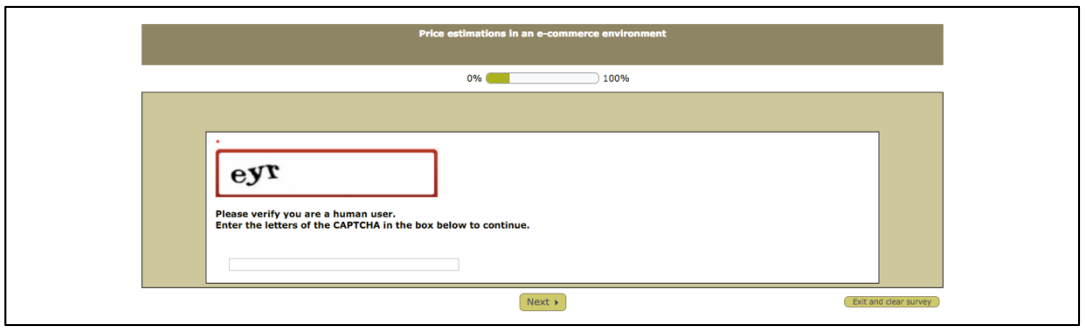




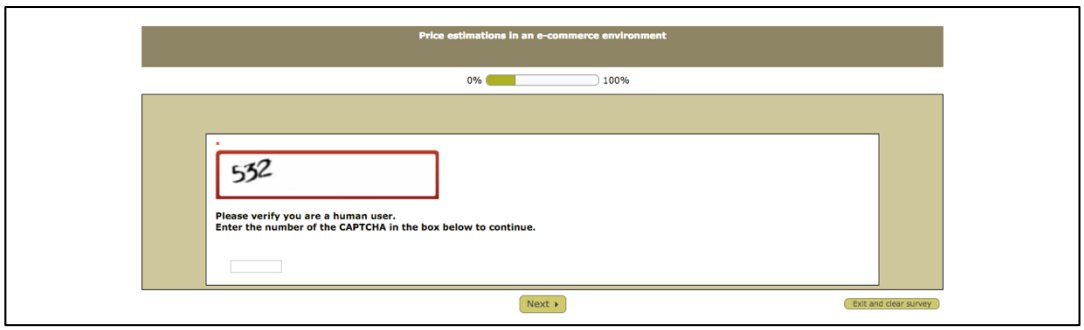


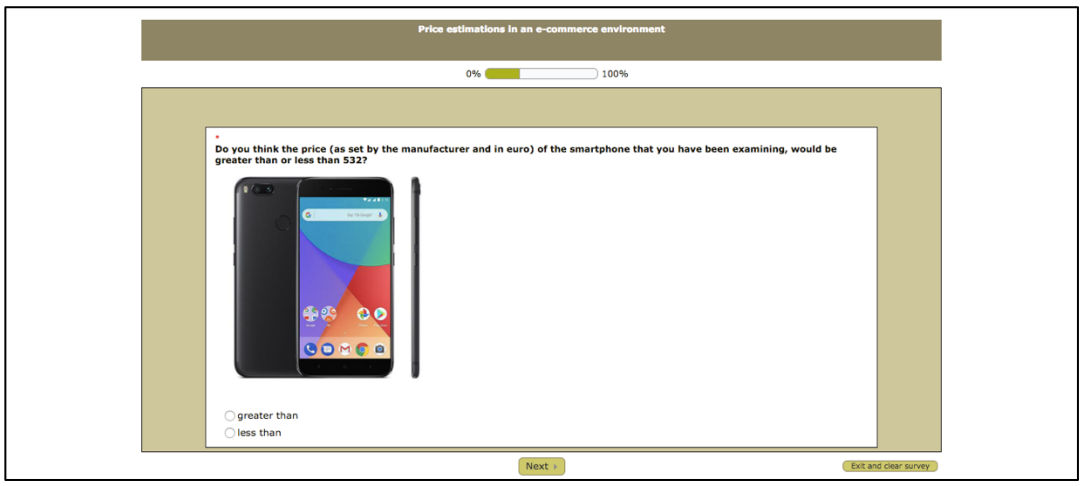


No anchor condition

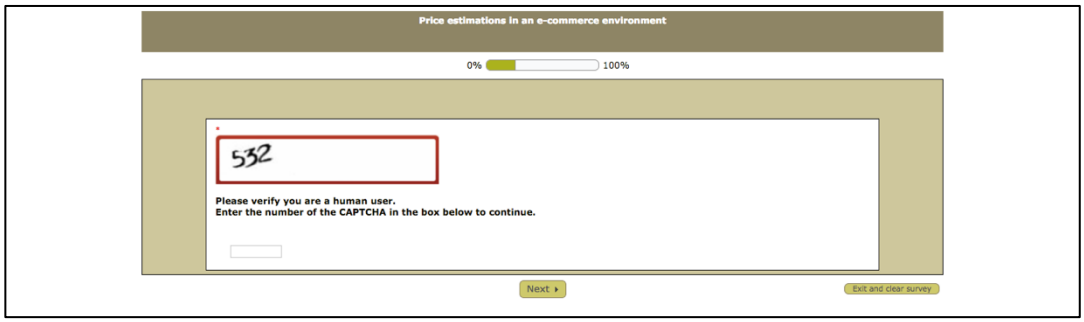


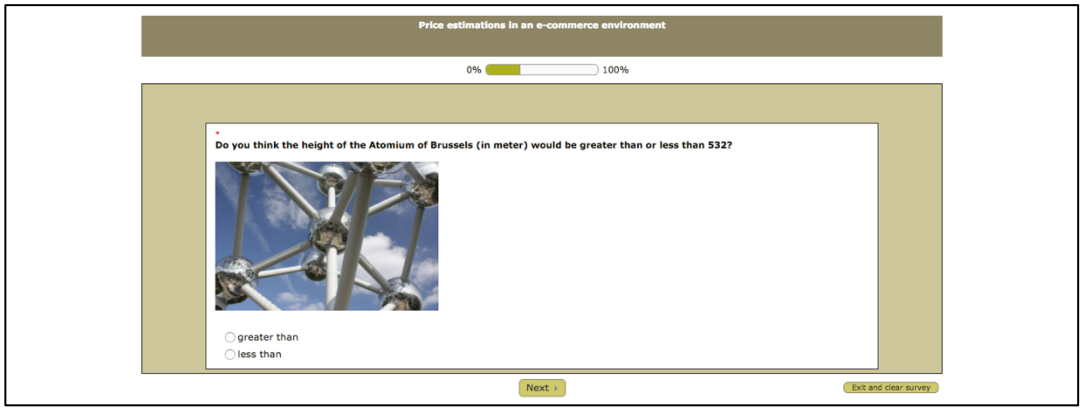
Standard anchoring condition



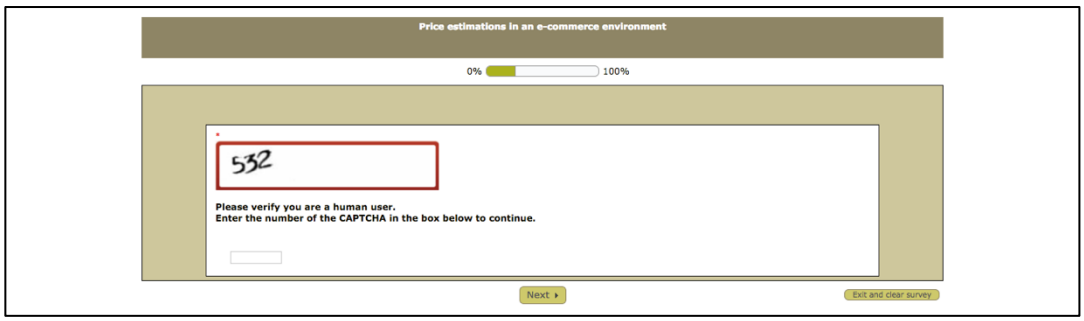


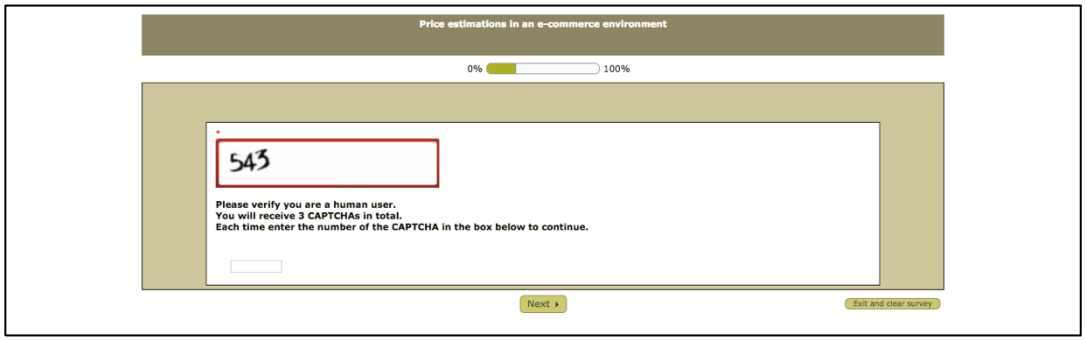
Unrelated comparison condition

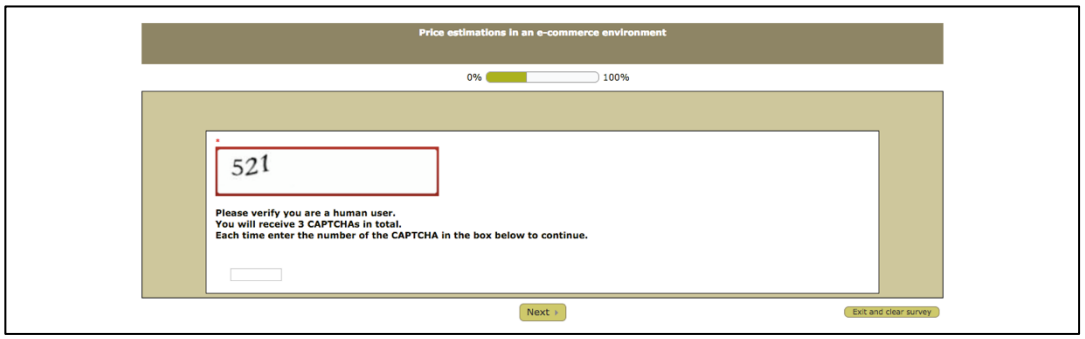




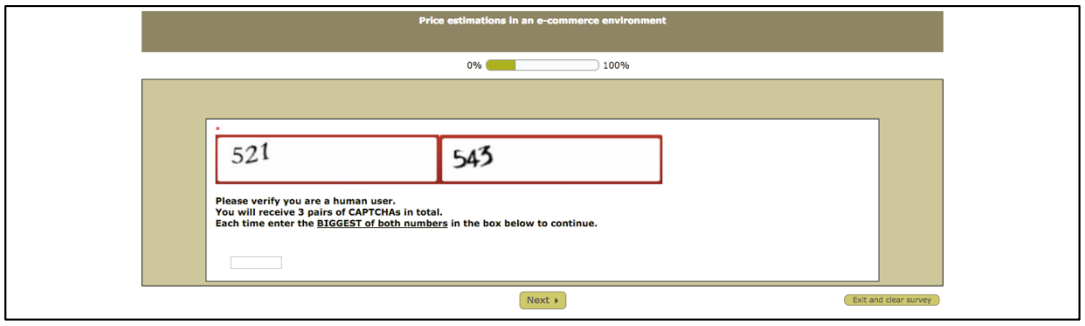
Copy condition

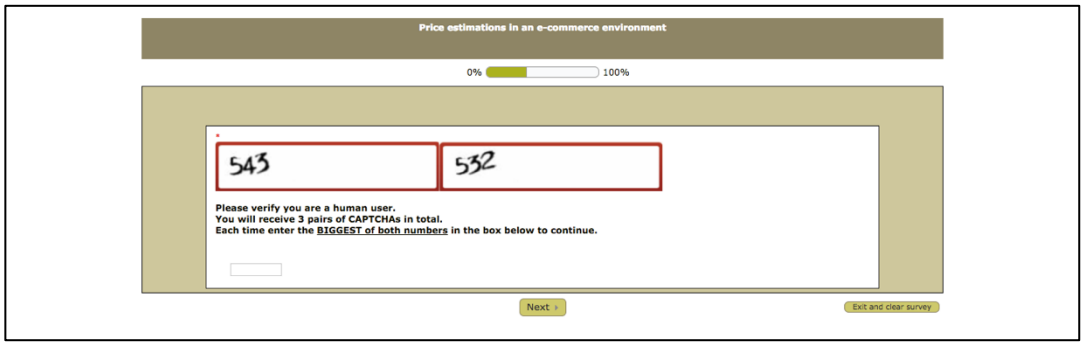


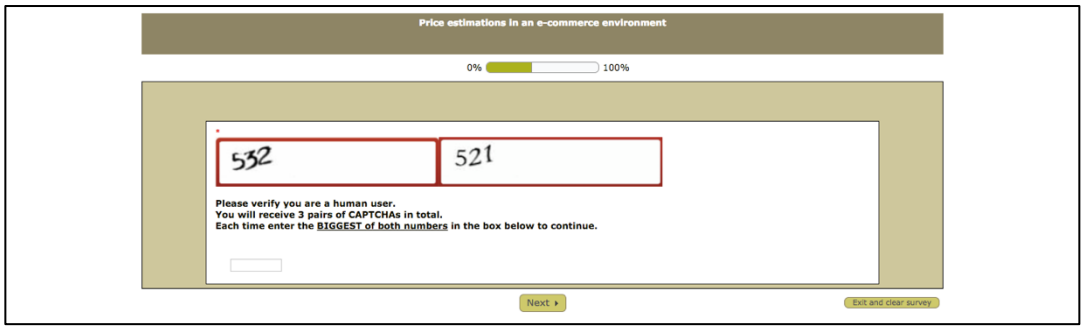




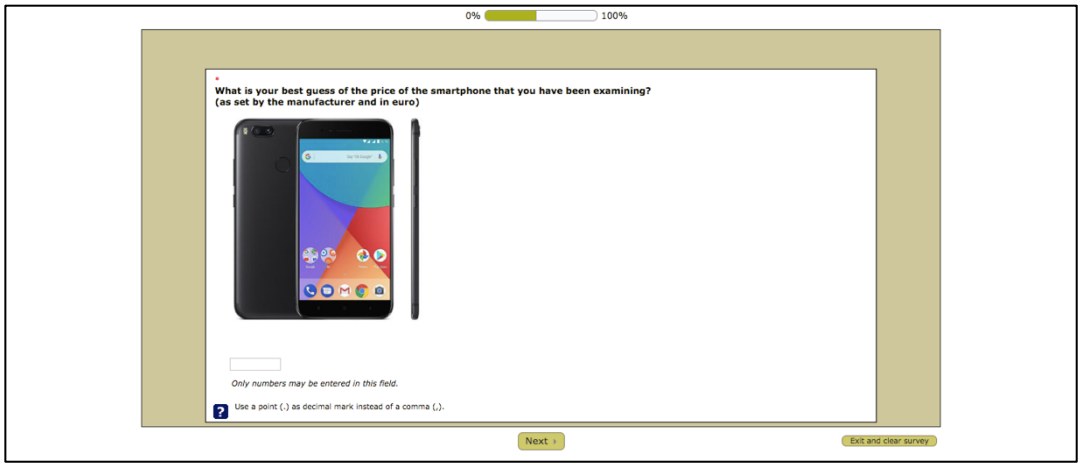
Computation condition

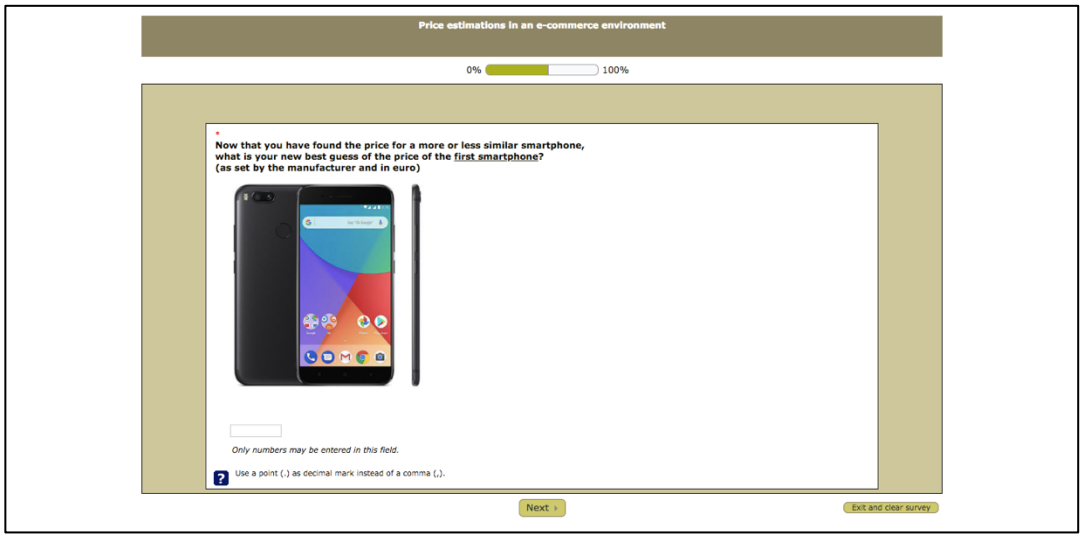


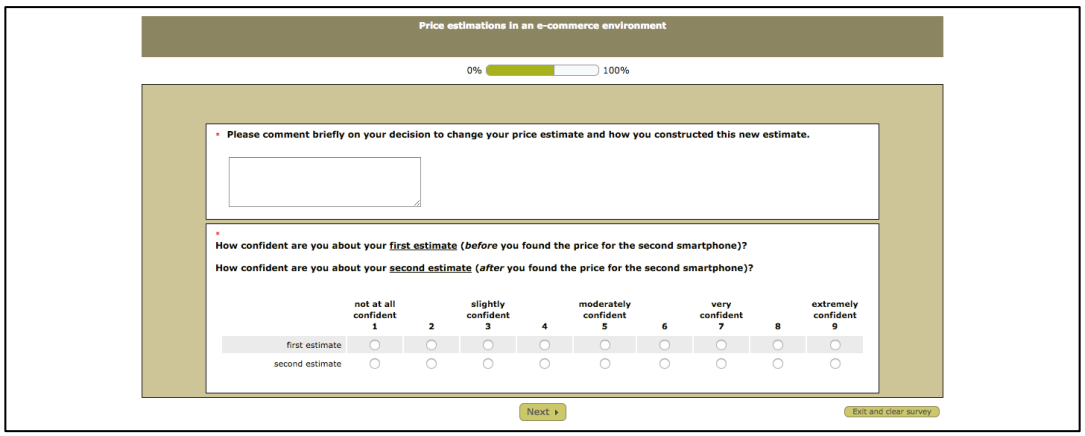


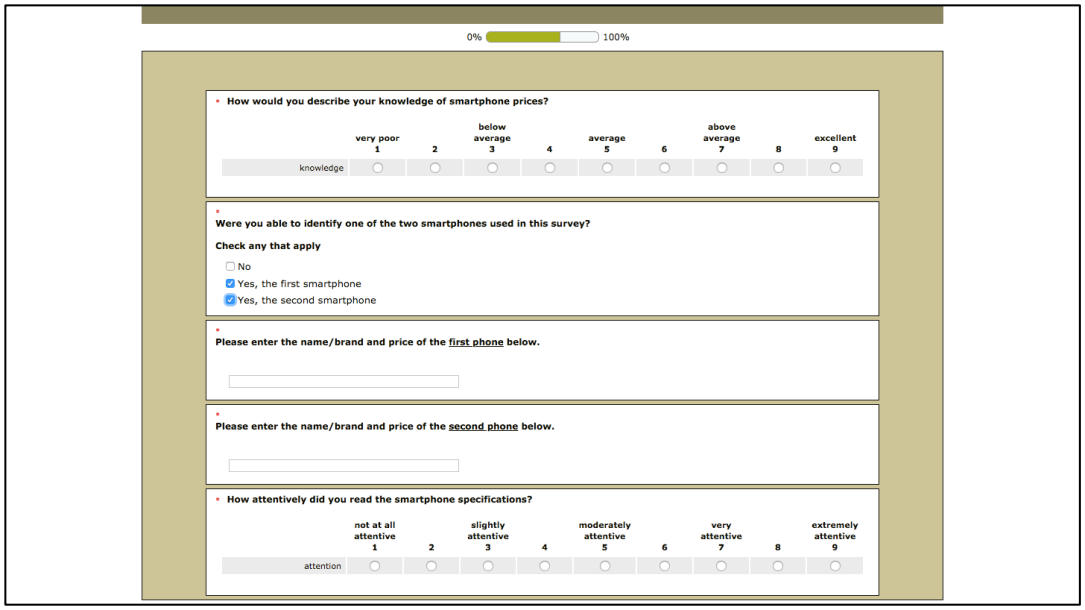


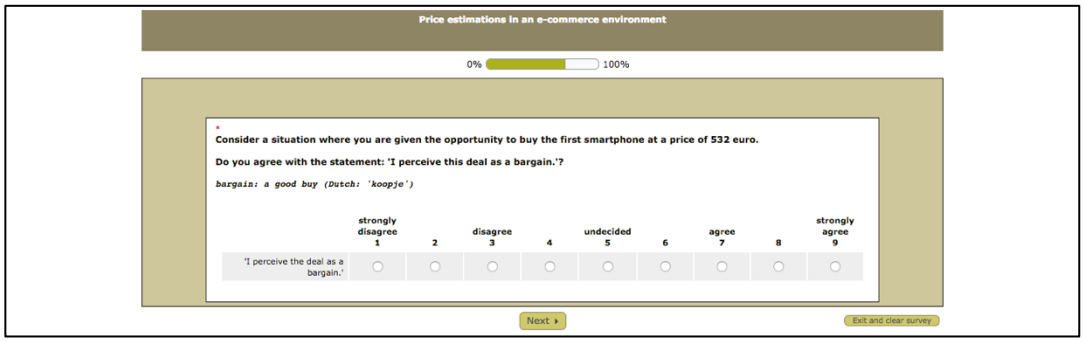
All conditions



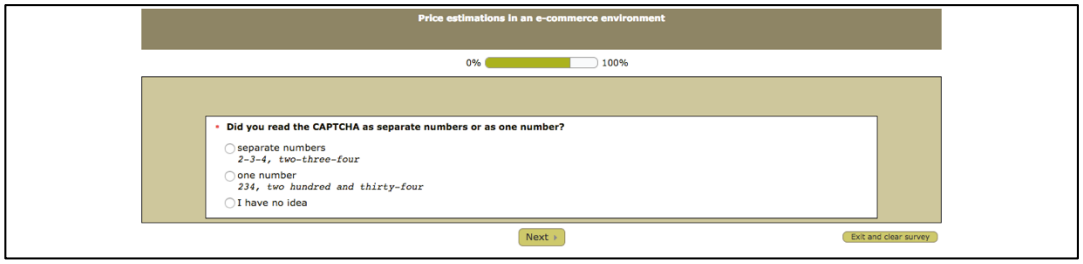

Price estimations in an e-commerce environment
0%
100%
Now that you have found the price for a more or less similar smartphone, what is your new best guess of the price of the first smartphone? (as set by the manufacturer and in euro)
Only numbers may be entered in this field.
Use a point (.) as decimal mark instead of a comma (,).
Next
Exit and clear survey


Price estimations in an e-commerce environment
0%
100%
Please comment briefly on your decision to change your price estimate and how you constructed this new estimate.
How confident are you about your first estimate (before you found the price for the second smartphone)?
How confident are you about your second estimate (after you found the price for the second smartphone)?
not at all confident 1
2
slightly confident 3
4
moderately confident 5
6
very confident 7
8
extremely confident 9
first estimate
second estimate
Next
Exit and clear survey

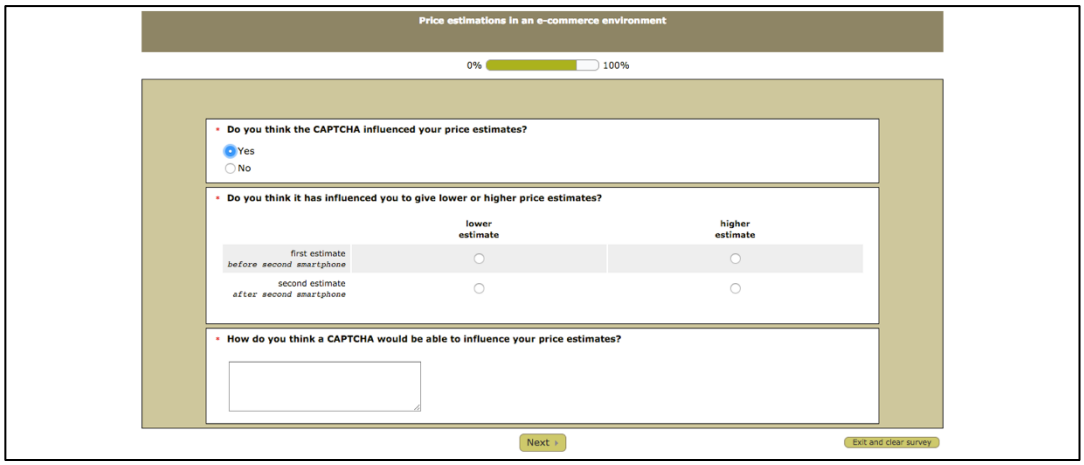

0%
100%
How would you describe your knowledge of smartphone prices?
very poor 1
2
below average 3
4
average 5
6
above average 7
8
excellent 9
knowledge
Were you able to identify one of the two smartphones used in this survey?
Check any that apply
No
Yes, the first smartphone
Yes, the second smartphone
Please enter the name/brand and price of the first phone below.
Please enter the name/brand and price of the second phone below.
How attentively did you read the smartphone specifications?
not at all attentive 1
2
slightly attentive 3
4
moderately attentive 5
6
very attentive 7
8
extremely attentive 9
attention

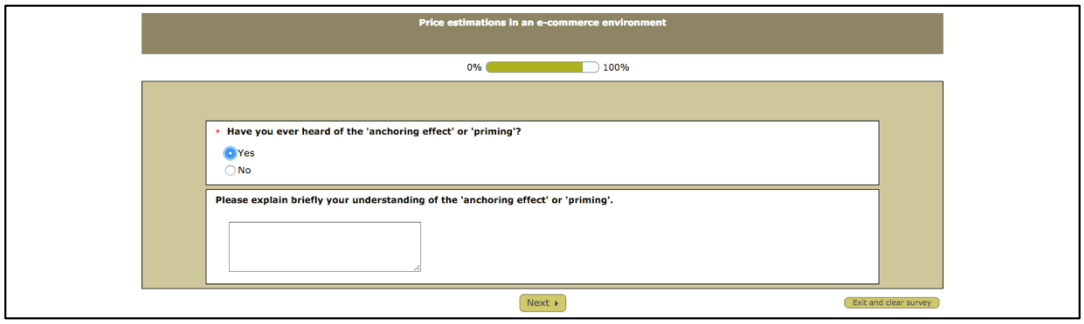


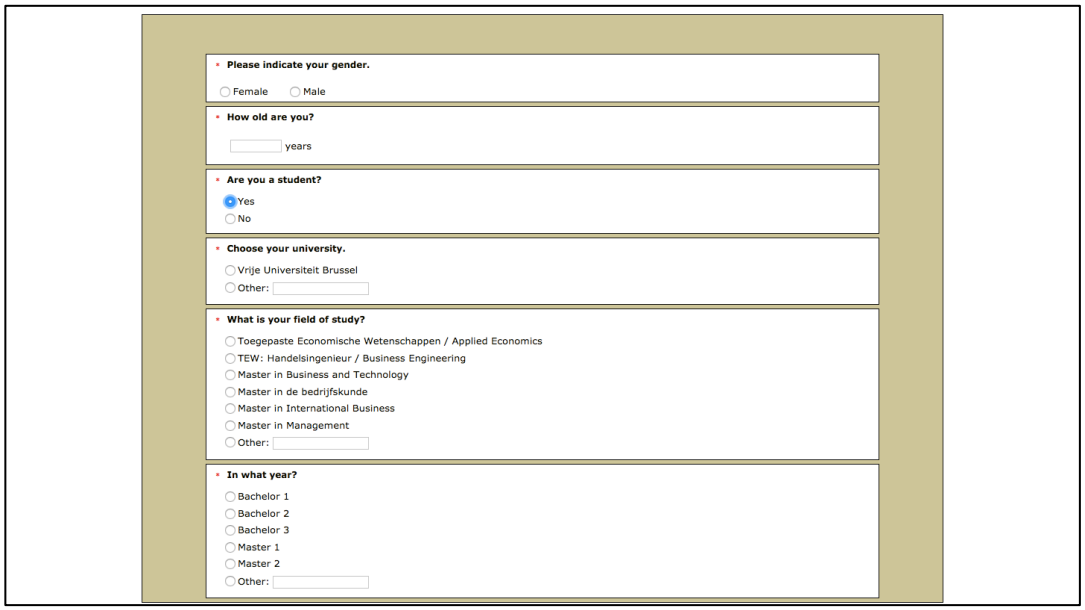
All conditions, except no anchor condition



All conditions







* **Please indicate your gender.**

○ Female ○ Male

* **How old are you?**

years

* **Are you a student?**

● Yes
○ No

* **Choose your university.**

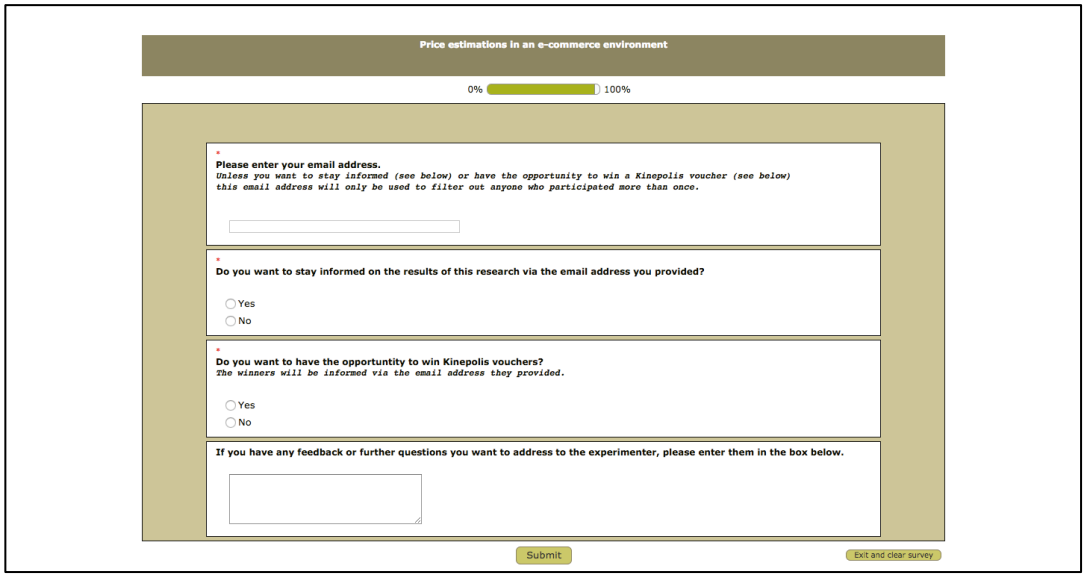
○ Vrije Universiteit Brussel
○ Other:

* **What is your field of study?**

○ Toegepaste Economische Wetenschappen / Applied Economics
○ TEW: Handelsingenieur / Business Engineering
○ Master in Business and Technology
○ Master in de bedrijfskunde
○ Master in International Business
○ Master in Management
○ Other:

* **In what year?**

○ Bachelor 1
○ Bachelor 2
○ Bachelor 3
○ Master 1
○ Master 2
○ Other:



Price estimations in an e-commerce environment

0% 100%

* **Please enter your email address.**
Unless you want to stay informed (see below) or have the opportunity to win a Kinepolis voucher (see below) this email address will only be used to filter out anyone who participated more than once.

* **Do you want to stay informed on the results of this research via the email address you provided?**

○ Yes
○ No

* **Do you want to have the opportuntity to win Kinepolis vouchers?**
The winners will be informed via the email address they provided.

○ Yes
○ No

If you have any feedback or further questions you want to address to the experimenter, please enter them in the box below.

Submit

Exit and clear survey

B Descriptive statistics of the participants

Table B.1

Descriptive statistics of the participants (N=30) in the calibration study

		participants
gender	male	12
	female	18
field of study	Applied Economics	21
	Business Engineering	6
	other	3
year of study	Bachelor 1	0
	Bachelor 2	21
	Bachelor 3	3
	Master 1	1
	Master 2	4

Note. The calibration study took place on March 8th, 2018, mainly during a lecture on 'Statistics II for Business Economics'.

Table B.2

Overview of the conducted experiments

date	lecture	participants
March 12th, 2018	International Monetary Economics	17
March 15th, 2018	Financial management and investment analysis	69
March 16th, 2018	International Monetary Economics	25
March 19th, 2018	Statistics I for Business Economics	72
March 19th, 2018	International Monetary Economics	20
March 20th, 2018	Information Systems Strategy and Management	29

Table B.3

Descriptive statistics of the participants (N=232)

		participants
gender	male	120
	female	112
field of study	Business Engineering	86
	Bedrijfskunde	67
	Applied Economics	60
	Master in International Business	12
	Master in Business and Technology	5
	other	2
year of study	Bachelor 1	68
	Bachelor 2	5
	Bachelor 3	59
	Master 1	76
	Master 2	21
	Manama	3
age	18 years	31
	19 years	25
	20 years	45
	21 years	35
	22 years	29
	23 years	37
	24 years	21
	25 years	9

C Impact of knowledge on the anchoring effect

The knowledge of the participants does not differ across conditions, since a Kruskal-Wallis test, shows no effect of the condition on the knowledge score ($\chi^2(4, N=183) = 1.30$, $p = .86$).

A two-way ANOVA (corrected for the unbalanced design with Type-III sums of squares) on the log price estimates revealed a significant main effect of experimental condition on the provided price estimate ($F(4, 173) = 2.38$ $p = .05$), and a significant main effect of the knowledge score on the estimate ($F(1, 173) = 3.63$, $p = .06$). This implies that the provided price estimate is influence by the knowledge of the participants[78]. There is however no significant interaction effect between the knowledge score and the experimental condition ($F(4, 173) = 1.21$, $p = .31$).

Finally, I performed a median split to see whether I could see anchoring effects similar to those in Table 8 in the subgroups after the median split. Table C.1 and Table C.2 give some information about this median split. If we look at the low knowledge group, a one-way ANOVA shows a significant effect of the experimental condition on the price estimates ($F(4, 71) = 2.08$, $p = .09$). However, a pairwise comparisons *t*-test shows no significant differences between specific conditions (p's >.15). For the high and median knowledge groups a Kruskal-Wallis test – the assumptions for ANOVA were not fulfilled – there are no differences between the mean estimates of the different experimental conditions ($\chi^2(4, N=49) = 2.84$, $p = .59$ and $\chi^2(4, N=58) = 2.52$, $p = .10$, respectively).

Conclusion

Knowledge has an impact on the price estimates of participants, but there is no evidence that the results in Table 8 can be explained by differences in the knowledge of the participants of the different experimental conditions, nor by a different effect of knowledge per experimental condition.

A median split on a bigger sample could provide more insight than the present 76 high-knowledgeable and 49 low-knowledgeable participants (divided over 5 conditions).

[78] There is a significant positive (Spearman's rank) correlation between the knowledge score and the price estimate ($r_s(181)=0.15$, $p = .04$). This is however not the case within the experimental conditions (p's >.35) except for the no anchor condition ($r_s(31)=0.34$, $p = .05$) and the unrelated comparison condition ($r_s(38)=0.32$, $p = .05$).

Table C.1

Number of participants in case of a median split

	low (1-4)	median (5)	high (6-9)	Total
no anchor	11	11	11	33
standard anchoring	19	6	12	37
unrelated comparison	14	16	10	40
copy	15	11	7	33
computation	17	14	9	40
total	**76**	**58**	**49**	**183**

Table C.2

Means of the price estimates per condition in case of a median split

	no anchor	standard anchoring	unrelated comparison	copy	computation
low (1-4)	315	458	383	369	453
median (5)	345	460	412	406	522
high (6-9)	407	424	473	377	516