

THINKING FAST AND SLOW ABOUT NUMBERS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Deepak Sirwani

August 2024

© 2024 Deepak Sirwani

THINKING FAST AND SLOW ABOUT NUMBERS

Deepak Sirwani, Ph.D.

Cornell University 2024

How do rating formats and time framing influence consumer perceptions and behaviors?

This dissertation explores these two important questions, focusing on the broader theme of numerical cognition and how the presentation of magnitudes—whether in the form of numerical ratings or time intervals—affects consumer judgment and decision-making.

Chapter 1 investigates how different formats of numerical ratings affect consumer behavior. Retailers use various rating formats, such as stars, circles, and Arabic numerals, to convey the quality and popularity of products and services. Through a series of ten studies, it is demonstrated that the format of these ratings affects consumers' perceptions of product quality, willingness to pay, and engagement with social media advertisements. Specifically, fractional ratings presented as graphical symbols (e.g., three and a half stars) are overestimated, while those presented in Arabic numerals (e.g., 3.5) are underestimated. This bias occurs because graphical ratings lead consumers to visually complete the image, anchoring their judgments on a rounded-up number, whereas Arabic numeral ratings anchor on the lower, left-most digit. These findings suggest that retailers should carefully consider their choice of rating formats to ensure accurate consumer perceptions and call for industry standards on rating formats.

Chapter 2 examines how time framing influences consumers' duration perceptions. In many product categories, such as whiskeys, vintage products, or second-hand goods, time

attribute plays an important role in determining the product's quality and value. Marketers often highlight the age of these products using different frames: a length frame (e.g., 12-years-old whiskey) or a boundary frame (e.g., 2012-whiskey). Through analyses of archival data and a series of experiments, it is found that time framed by its length (e.g., "10 years later/ago") is perceived as more distant compared to year framing (e.g., "2034 / 2014"), which ultimately influences product's perceived value. This research advances the literature on numerical cognition by revealing how nominal values used in different time frames influence consumer perceptions and behaviors. This effect of time framing has critical implications for how consumers evaluate products with a time component, weigh time against money, and make investment decisions.

In conclusion, this dissertation advances knowledge in two areas: the impact of rating formats on consumer behavior and the effect of time framing on duration perceptions. Both projects underscore the importance of understanding how numerical cognition—specifically, the presentation of magnitudes—affects consumer judgments. This theme of thinking fast and slow about numbers highlights the cognitive processes that lead to biases in perception, providing valuable insights for marketers, retailers, and researchers in developing more effective communication strategies and enhancing consumer decision-making processes.

BIOGRAPHICAL SKETCH

Deepak Sirwani was born in Jaipur, India. He received his BTech in Electrical Engineering from IIT Kharagpur and his Master and Ph.D. in Marketing from Cornell University. His research focuses on consumer judgments and decisions, particularly within digital environments. Before academia, he was a founding team member of a global technology company and worked as an Intellectual Property Law Consultant for US law firms. His research interests include numerical cognition, time perception, and decision-making broadly.

I dedicate this dissertation to the late Daniel Kahneman, whose book "Thinking, Fast and Slow" inspired me to leave my job and pursue a Ph.D. in human decision-making. I also extend my deepest gratitude to my friends and family for their unwavering support during this seemingly radical journey.

ACKNOWLEDGMENTS

Reflecting on the incredible journey of the past six years, I am filled with gratitude for the support and encouragement that made this dissertation possible. First and foremost, I would like to thank my advisor, Manoj Thomas. When I moved to Ithaca from India, I had little idea about research or what living in the US would be like. The first four years were very challenging—figuring out coursework, dealing with culture shock, and learning just how tough research can be. Manoj, you stood by my side even though the initial years were met with constant failures in research. Going through these struggles made the entire journey more meaningful. Your guidance and mentorship have been invaluable, and I hope to make you proud in the long term.

I am deeply grateful to Suzanne Shu, whose unwavering belief in me and constant encouragement were a source of strength. Your reassuring words and support have greatly impacted my journey. Special thanks also to Stijn van Osselaer for being a guiding light with your insights and feedback. I would also like to express my gratitude to my committee members, Laura Neimi and Felix Thoemmes. Laura, thank you for always supporting me. Felix, thank you for teaching me statistical skills and being patient with my constant inquiries.

I am thankful for the wonderful friendships I have made, particularly with those who shared the struggles of this journey. Your companionship made this experience more meaningful. A heartfelt thanks to my friend Rahul Naryani, who always helped me brainstorm research ideas, and to my brother Kartik Sirwani, who assisted with programming needs from time to time. Your support has been crucial.

Lastly, I extend my deepest gratitude to my friends and family for their unwavering support. This dissertation is as much yours as it is mine.

TABLE OF CONTENTS

Biographical Sketch	v
Dedication	vi
Acknowledgments	vii
Chapter 1	1
Chapter 2	54
Chapter 1 Appendix	111
Chapter 2 Appendix	141

CHAPTER 1

FOOLED BY STARS:

MISESTIMATIONS IN JUDGMENTS OF NUMERIC RATINGS

Abstract

Retailers use different formats of numerical ratings (e.g., stars, circles, Arabic numerals) to communicate the quality and popularity of products and services. Do consumers' evaluations of ratings communicated using stars differ from those communicated in Arabic numerals? Results from ten studies demonstrate that rating format influences consumers' perceptions of product quality, willingness to pay, and engagement with social media advertisements. The rating format influences consumers' responses to fractional ratings, but it does not influence their responses to whole number ratings. Consumers overestimate fractional ratings presented as graphical symbols (e.g., three and a half stars) and underestimate those presented in Arabic numerals (e.g., 3.5). The overestimation in graphical ratings is caused by the visual completion of incomplete images, leading consumers to anchor their judgments on rounded-up numbers. That is, the evaluation of three and a half stars is anchored on four stars. In contrast, the underestimation of Arabic numeral ratings is caused by left-digit anchoring, leading consumers to anchor on rounded-down numbers. That is, the evaluation of 3.5 is anchored on 3. These findings imply that retailers should judiciously select rating formats. Moreover, these findings call for industry standards on rating formats.

The use of numerical ratings to make evaluative judgments has become increasingly prevalent. Most commercial and social interaction platforms incorporate some form of numeric rating system. Numerical ratings are used to evaluate books, movies, and products on commercial platforms; social communications on social media platforms; places to visit on navigation platforms; and even people on dating and relationship facilitating platforms. It has been suggested that product rating is the most significant factor impacting purchase decisions, even more important than factors like price, brand, and recommendations from friends and family (PowerReviews 2021). According to a recent report from McKinsey & Company, even a small increase in numerical ratings, such as 0.2, can lead to a significant increase in product sales, ranging from 30-200% (McKinsey 2021).

Although the use of ratings is ubiquitous, there is considerable heterogeneity in the format of ratings. Different platforms use different symbols to communicate the ratings. The two most popular forms of rating formats are graphic symbols, such as stars or circles, and Arabic numerals. Some platforms, such as Amazon, Yelp, and Tripadvisor, use graphic symbols such as stars or circles to represent ratings (e.g., an image of three and a half stars). Others, such as Uber, Facebook, and Airbnb, use Arabic numerals (e.g., 3.5). And some, such as Google Maps, Walmart, and Goodreads, use a combination of both types of symbols—stars and numerals together—to represent ratings.¹ In this research, we examine how the format of the ratings influences evaluative judgments. Do consumers' evaluations of ratings communicated using stars differ from those communicated in Arabic numerals? When and why do rating formats influence the perceived magnitude of the ratings? Given the lack of an industry standard for rating

¹ See Table 1 in the Chapter 1 Appendix for more examples of retailers using graphical, Arabic numeral, or a combination of both rating symbols.

symbols, the answers to these questions will assist retailers in making a more effective choice of rating symbol.

Surprisingly, there is limited research on how different rating symbols impact rating perception. A notable exception is the recent paper by Jia, Wan, and Zheng (2023). These authors compare the effects of using stars and bars on perceptions of product ratings. The present paper differs from Jia et al.'s work in important ways. We compare perceptions of ratings presented using stars and Arabic numerals. Retailers must often decide whether to use stars, Arabic numerals, or a combination to communicate ratings.¹ Our results can inform this important decision, while the previous work cannot. Additionally, while previous work studies only relative perceptions of ratings using stars and bars, as it will become clear later, we use a number line estimation task that is designed to compare both relative magnitude perceptions (i.e., relative magnitude perceptions of stars and Arabic numerals) as well as the accuracy of each magnitude perception (whether star ratings are underestimated or overestimated, and whether Arabic numerals are underestimated or overestimated). This is the first research to study how retailers' use of stars instead of Arabic numerals to represent product ratings can influence consumer judgments. This research, to our knowledge, is the first to apply the number line estimation task from cognitive psychology literature (Booth and Siegler 2008; Dehaene et al. 2008; Schley and Peters 2014; Siegler and Opfer 2003) to examine how rating symbols affect the accuracy of magnitude perceptions in product ratings.

We report results from eight controlled experiments and two experiments on a social media platform that examine the effects of rating format on consumers' evaluative judgments of ratings. In our experiments, we presented the same ratings either using graphic symbols (e.g., image of three and a half stars) or using Arabic numerals (e.g., 3.5). To presage our results, we

find that people tend to overestimate the magnitude of star ratings and underestimate that of Arabic numeral ratings. Furthermore, the misestimation of ratings is only observed for fractional ratings (e.g., 3.25, 3.5, 3.75, etc.), not for round ratings (e.g., 2.0, 3.0, 4.0, etc.). This result rules out the possibility that the misestimation is caused by positivity associated with stars or other graphical imagery. Our experiments were designed to characterize the perceptual and cognitive processes that underlie these misestimations in magnitude judgments of ratings. Our findings show that the type of rating format used can influence the perceptual and cognitive phenomena in rating magnitude judgments. When graphic symbols are used, the magnitude judgment is influenced by visual completion, leading to an overestimation of ratings. When Arabic numerals are used, the magnitude judgment is affected by left-digit anchoring, resulting in an underestimation of ratings.

Our results have important implications for retailers, suggesting that they should judiciously choose the type of rating symbols they use. The findings also call for government regulation and industry standards on product rating formats. The lack of industry standards can make it difficult for consumers to meaningfully use product ratings to inform their purchasing decisions.

Additionally, our findings also have implications for theory, intersecting and contributing to three distinct research streams in consumer behavior: visual perception, product ratings, and numerical cognition. Although there is a rich stream of research documenting how visual perception influences judgments (Bagchi and Cheema 2013; Barasz et al. 2017; Hagtvedt and Brasel 2017; Krishna 2006; Raghurir and Krishna 1999; Townsend and Kahn 2014), the literature on product ratings and numerical cognition have largely ignored the role of perceptual processes in consumers' reactions to different rating symbols. With the exception of Jia et al.

(2023), most previous research on product ratings has centered around the informational aspects of the ratings, such as average rating (Chen and Lurie 2013; Chevalier and Mayzlin 2006; de Langhe, Fernbach, and Lichtenstein 2016a; b), number of ratings (Watson, Ghosh, and Trusov 2018), rating variance (Fisher, Newman, and Dhar 2018; Rozenkrants, Wheeler, and Shiv 2017; Schoenmueller, Netzer, and Stahl 2020), and rating scales (Kyung, Thomas, and Krishna 2017).

In a similar vein, while a rich stream of research has examined how numerical cognition affects various aspects of consumer behavior, including magnitude judgments (Bagchi and Li 2011; Monga and Bagchi 2012; and Pandelaere, Briers, and Lembregts 2011) and consumer preferences (King and Janiszewski 2011; Lembregts and Pena-Marin 2021; Yan and Sengupta 2021), the role of perceptual symbols in such judgments has been largely ignored. A case in point, conceptual frameworks documented in recent review articles on numerical cognition (Thomas and Morwitz 2009; Santana, Thomas, and Morwitz 2020) are not helpful in explaining why different perceptual symbols, such as star versus Arabic numerals, might have divergent effects on numerical cognition. Our work intersects these three fields, highlighting the role of visual processing of symbols in numerical cognition and product ratings. We present a conceptual framework that can be used to explain how visual processing of symbols influences the magnitude perception of fractional numbers, predicting when fractional numbers will be overestimated or underestimated.

THEORETICAL BACKGROUND

Our conceptualization of the impact of symbols on numerical magnitude judgments is based on two postulates. First, because processing fractional numbers can be challenging (relative to whole numbers), people tend to rely on heuristics for magnitude judgments.

Specifically, we propose that while judging the magnitude of fractional numbers, people use the nearby whole numbers as initial anchors.

Second, these anchors vary depending on the type of symbol used. In the case of graphical ratings, such as stars, the automatic visual completion of fractional stars leads people to anchor their judgments on rounded-up numbers, thus resulting in overestimation. For instance, the evaluation of three and a half stars is anchored on four stars. In contrast, in the case of Arabic numeral ratings people anchor on the left-most digit, leading people to anchor on rounded-down numbers, thus resulting in underestimation. That is, the evaluation of 3.5 is anchored on 3.0. In the following sections, we will first review prior research on numerical cognition and visual perception that lay the foundation for our hypotheses.

Processing Fractional Quantities

The analog model of numerical cognition posits that people make sense of numeric symbols by converting the symbols into internal approximate magnitudes along a mental number line, a process known as mapping of symbolic representation to analog magnitudes (Dehaene 1997; Dehaene, Dupoux, and Mehler 1990; Hinrichs, Yurko, and Hu 1981; Schley and Peters 2014). Inaccurate mapping can lead to biases in magnitude judgments.

It is easy for humans to instinctively understand the magnitude of whole numbers, such as 1, 2, and 3, because they encounter these types of numbers more frequently and have evolved to instinctively judge their values. That is, people can easily map whole numbers such as 1, 2, and 3 on their mental number line that they use for magnitude judgments. However, this is not true for fractions. Although children are able to count from a young age, they do not understand fractions until later in their development (Dehaene 1997). Even adults struggle to assess the magnitudes of fractional numbers, such as 3.2, 3.5, or 3.8 (Gigerenzer and Hoffrage 1999).

When they encounter such fractional numbers, salient round numbers serve as initial anchors that help them judge the magnitude of the fractional numbers (see Rosch 1975 for more on round numbers as cognitive reference points in numerical judgments). That is, when mapping a fractional number onto the internal mental number line, the human mind often starts with a salient round number as an anchor and adjusts from there. However, the adjustments in anchoring-and-adjustment heuristics are usually insufficient (Epley and Gilovich 2004; Griffin and Tversky 1992; Tversky and Kahneman 1974), leading to estimates that may be lower or higher than the actual number, depending on whether the chosen anchor is below or above the actual number.

Anchors Differ by Rating Symbols

We posit that rating symbols can influence the choice of the initial anchor. When presented with graphical symbols, people tend to anchor on rounded-up numbers, leading to an overestimation of the magnitude of fractional ratings. However, when presented with Arabic numerals, people are likely to anchor on rounded-down numbers, leading to an underestimation of the magnitude of fractional ratings. This systematic proclivity to choose rounded-up (vs. rounded-down) anchors during the internal representation of fractional graphical (vs. Arabic numeral) ratings is influenced by two distinct visual processes.

Visual Completion Bias. Magnitude judgments of fractional star ratings are influenced by visual completion of incomplete images, which leads to the overestimation of fractional graphical ratings. Visual completion is the psychological process by which the brain uses contextual clues, prior knowledge, and expectations to fill in gaps in sensory information and create a coherent perception of the world around us (Kanizsa 1979; van Lier and Gerbino 2015; Pessoa and De Weerd 2003; Pessoa, Thompson, and Noë 1998). This process is a fundamental

aspect of human perception as it allows us to perceive objects and scenes as whole and complete, even when our sensory information is incomplete or degraded (Ramachandran 1992). Our innate need for visual completion leads us to perceive images as complete, even when they are not (Foley et al. 1997, 2007; Kanizsa 1979; Pessoa, Thompson, and Noë 1998). Research has shown that visual completion is a rapid, automatic process that occurs at early stages of vision (Rensink and Enns 1998). It involves removing occlusion edges and linking fragments together, allowing us to perceive objects as whole and complete. The completed structures, and not the fragments, then become the units that subsequent recognition processes work with (Rensink and Enns 1998).

Figure 1.1: Initial Anchors Elicited by Fractional Ratings



Fractional ratings shown using graphical symbols, such as an image of three and a half stars, are visually incomplete, and our brains automatically try to fill in the gaps. For example, when people see the image of three and a half stars, the fourth star is incomplete as it is only half-filled. In the early stages of perception, visual completion causes people to perceive the

image of three and a half stars as four complete stars, leading to an initial anchor of 4.0 for the magnitude judgment. Even when people subsequently correct their initial perception, these initial anchors surreptitiously influence their magnitude judgments. This results in an overestimation of the fractional star ratings. See Figure 1.1 for a schematic depiction of the postulated mechanism.

Left-Digit Bias. Fractional ratings using Arabic numerals, in contrast, are biased by left-digit anchoring, which leads to the underestimation of fractional Arabic numeral ratings. Research has shown that people tend to focus heavily on the left-most digit when processing numbers written in Arabic numerals (Thomas and Morwitz 2005). This phenomenon leads people to anchor on the left-digit, i.e., rounded-down numbers in the case of fractional numerical ratings. One of the proposed reasons for this bias is the way the human mind reads numbers, which is typically from left to right. While reading multi-digit numbers, the encoding process begins with the leftmost digit and then proceeds to the right. For example, when presented with the number “3.5”, people are more likely to first attend to “3” and anchor their judgment of the number on this digit, leading to an underestimation of the actual magnitude. Previous research has found that the left-digit bias can lead to significant biases in judgments and decisions in consequential settings including stock market transactions and public utility evaluations (Bhattacharya, Holden, and Jacobsen 2012; Ginzberg 1936; Jiang 2022; Lacetera, Pope, and Sydnor 2012; Manning and Sprott 2009; and Stiving and Winer 1997). For example, one study showed that even small changes in a school’s numerical average grades that changed the leftmost digit of the grade resulted in large shifts in people’s evaluation of the school’s performance (Olsen 2013).

HYPOTHESES

Our conceptualization of how rating format influences magnitude perception yields

several novel predictions. We hypothesize that star ratings will activate rounded-up numbers as anchors because of visual completion (e.g., four will be the anchor for three and a half stars). In contrast, Arabic numeral ratings will activate the left-most digits as anchors, which will be the rounded-down numbers (e.g., 3 will be the anchor for 3.5). Thus, fractional ratings will be misestimated depending on the rating format, such that:

H1a: Fractional ratings presented using graphical symbols (e.g., three and a half stars) will be perceived to be higher than the actual rating.

H1b: Fractional ratings presented in Arabic numerals (e.g., 3.5) will be perceived to be lower than the actual rating.

Importantly, we predict that these systematic misestimations in evaluating consumer ratings will have significant downstream effects on consumer behavior. This includes influencing perceptions of product quality, affecting consumers' willingness to pay for these products, and even changing their engagement with social media advertisements related to these products.

Boundary Conditions

Our conceptualization of the effect of rating format on perceived magnitude identifies an important boundary condition. It predicts that visual completion will increase the perceived magnitude of fractional star ratings (e.g., three and half stars), but not whole-numbered star ratings (e.g., four full stars), since there is no scope for visual completion for the latter. Similarly, we expect that left-digit bias will reduce the perceived magnitude of fractional Arabic numerals (e.g., 3.50), but not whole numbered Arabic numerals (e.g., 4.00), as the left digits of the latter do not elicit a lower anchor. Thus,

H2: The overestimation of graphical ratings and the underestimation of Arabic numeral ratings will manifest for fractional ratings (e.g., 3.25, 3.50, 3.75) but will be attenuated for whole number ratings (e.g., 3.00, 4.00).

It is worth clarifying that this boundary condition does not limit the scope or the impact of the proposed effect, as a large number of products sold on e-commerce platforms tend to have fractional average ratings. For instance, in an examination of the snack category on Amazon USA, we observed that approximately 80% of snack products (out of 8,637 items) had an average fractional rating. This figure rose to 92% when we filtered products with at least ten ratings (totaling 6,234 items).²

Our conceptualization also identifies another boundary condition, namely, the visual completeness of the images used in graphical ratings. The root cause of the overestimation of graphical ratings is the automatic propensity to complete visually incomplete pictures. It is possible to alleviate this by creating graphical ratings that do not seem visually incomplete; see Study 3 stimuli in Figure 1.2. Thus, if we use visually complete images to depict the fractional ratings, the perceptual system will be less likely to complete the graphical symbols. This should attenuate the overestimation of graphical ratings. Formally,

H3: The overestimation of graphical ratings will be attenuated if the fractional symbols are visually complete.

Note that since visual completion is an automatic process, we do not expect this intervention to completely remove the overestimation; we expect it to reduce the extent of overestimation.

² See Figure 1 in Chapter 1 Appendix for details on the distribution of these average ratings.

Figure 1.2: Rating Symbols Used in All Studies



Figure 1.2—This figure shows the symbols used for numeric ratings. In graphic ratings, the magnitude of the rating is represented by the percentage of the area that is colored unless specified that linear proportional fill is used. For example, a rating of 3.25 would be represented by three fully colored shapes and a fourth shape with its area 25% filled. All stimuli were designed with AutoCAD software, while those used in Supplementary Study 3 with linear proportion were sourced from Jia, Wan, and Zheng (2023).

Distance from Anchors

Finally, our theory suggests that the pattern of misestimation will vary for star and Arabic numeral ratings. Fractional star ratings are overestimated because of anchoring on the higher round numbers, and thus, we predict that the smaller fractions will experience greater overestimation. For example, quarter star ratings (e.g., 3.25) will be overestimated more than three-fourth star ratings (e.g., 3.75). In contrast, fractional Arabic numerals are underestimated due to anchoring on the lower round numbers, leading to greater underestimation for larger fractions. That is, three-fourth ratings (e.g., 3.75) will be underestimated more than quarter ratings (e.g., 3.25).

These predictions stem from the anchoring and adjustment literature, as demonstrated in Epley and Gilovich's work (2004, 2006), which posits that the insufficiency of the adjustment increases as the distance between the anchor and the stimulus increases. This is because of the iterative nature of the anchoring and adjustment process. In a magnitude estimation task, people begin by making small adjustments from the initial anchor and evaluate the plausibility of the adjusted estimate. If the estimate appears plausible, the adjustment halts; otherwise, further adjustments are made in small steps until a plausible value is reached. As Epley and Gilovich (2006) empirically demonstrated, the adjustments are typically insufficient. Furthermore, the larger the gap between the anchor and the stimulus, the greater the insufficiency of the adjustment is. This literature suggests that the error in estimations of fractional ratings will be in the direction of the anchor and proportional to the distance between the anchor and the rating.

In the case of star ratings, people will start with rounded-up anchors and then adjust downwards. For example, when they see the 3.25 stars and 3.75 stars, they will anchor their judgment on 4 stars, and start adjusting downwards. The adjustment will be less sufficient for

smaller fractions (e.g., 3.25 stars) than for larger fractions (3.75 stars). In contrast, when these same ratings are presented as Arabic numerals, they will anchor their judgments on the lower round number 3 and start adjusting upwards. The adjustment will be less sufficient for larger fractional numbers (3.75) than for smaller numbers (3.25). Formally, the degree of misestimation for fractional ratings will be proportional to the distance between the anchor and the fractional rating, such that:

H4a: For graphical ratings, the overestimation will be greater for quarter ratings (e.g., 3.25) than for three-fourth ratings (e.g., 3.75).

H4b: For Arabic numerals, the underestimation will be greater for three-fourth ratings (e.g., 3.75) than for quarter ratings (e.g., 3.25).

Alternate Accounts

To explain the overestimation of star ratings, we considered several plausible alternative mechanisms beyond the visual completion of incomplete stars. One such mechanism is the cognitive tendency to round fractional star ratings to the nearest whole number due to mental shortcuts or cognitive miserliness. This rounding could explain the overestimation of star ratings above the half mark (e.g., 3.75), but it does not explain overestimations for star ratings below the half mark (e.g., 3.25). Another account is that consumers may overestimate fractional star ratings due to the positive associations with star shapes, inadvertently influencing their magnitude judgments. However, per this theory, we should not see the effect for non-star graphical symbols like circles or squares. Additionally, per this account, we should see overestimation for complete stars, not just fractional stars.

A third consideration is the uncertainty inherent in perceiving incomplete graphical symbols. For instance, with an image of three and a quarter stars, it might be unclear how much

of the fourth star is filled in, as opposed to the unambiguous representation of 3.25 in Arabic numerals. While this perceptual uncertainty could lead to more erroneous judgments, it does not inherently explain the systematic overestimation observed. Our empirical studies delve into these alternative explanations, assessing their validity in the context of our findings.

We used ten studies ($N \sim 3,000$ for lab experiments and $N \sim 45,000$ for field experiments) to test our predictions. All studies reported in the paper were pre-registered on the Open Science Framework, and the materials, surveys, raw data, and R codes used in the experiments are publicly accessible online (<https://bit.ly/3YgBJs1>).

STUDY 1: NUMBER LINE ESTIMATION

Study 1 was designed to investigate whether the type of symbols used to represent numerical quantities can influence perceptions of magnitude. For this study, we employed a mental number line task, commonly used in numerical cognition research to examine intuitive magnitude judgments. The mental number line task is a widely accepted method to measure people's intuitive magnitude judgments (Booth and Siegler 2008; Dehaene et al. 2008; Schley and Peters 2014; Siegler and Opfer 2003). We adapted this task to test our hypotheses. Participants were shown several product ratings and asked to indicate their perceived magnitude on a number line. The ratings were presented either as stars, Arabic numerals, or both stars and Arabic numerals (see Study 1 stimuli in Figure 1.2). Thus, we used the number line task to test H1, H2, and H4, along with examining the effect of multiple representations (i.e., star and Arabic numerals together) on the subjective magnitude perception. The number line task not only allows us to test the relative magnitude perceptions of star and Arabic numeral ratings, but also allows us to test whether each rating is underestimated or overestimated on the number line. This study

was pre-registered at OSF (<http://bit.ly/3Wtri2k>).

Participants and Procedure

We analyzed data from 616 participants (44.7% female, 54.3% male, 1% nonbinary, $M_{\text{age}} = 41$ years) who were part of a U.S. standard sample from Prolific. For more detailed information regarding data exclusions, please see the Chapter 1 Appendix. Participants were randomly assigned to one of three conditions that varied in the type of symbols used: star ratings, Arabic numeral ratings, or both star and Arabic numeral ratings. All participants estimated the position of several ratings on an unmarked horizontal line with endpoints 1 and 5. This task allowed us to map how participants encode the magnitude of the ratings on the mental number line. Before administering the mental number line task, participants were put through a calibration phase; they were trained to correctly identify the position of three numeric ratings—1, 3, and 5—on a marked horizontal line with markers at 1, 2, 3, 4, and 5 (see Chapter 1 Appendix for details). The numeric ratings in the calibration phase and in the mental number line task were shown using rating symbols based on the assigned rating symbols condition; participants in the star ratings condition saw the ratings represented as stars, those in the Arabic numeral condition saw the ratings as Arabic numerals, and those in the third condition saw both star and Arabic numeral ratings (see Study 1 stimuli in Figure 1.2).

In the mental number line task, participants were presented with 17 numeric ratings from 1 to 5 in increments of 0.25, one at a time. The order of the 17 judgments was randomized for each participant. They were asked to estimate the position of each rating on an unmarked horizontal line with endpoints 1 and 5 (see Chapter 1 Appendix for a picture). The ratings included five whole numbers (1.00, 2.00, 3.00, 4.00, 5.00), four fractional numbers with quarters

(1.25, 2.25, 3.25, 4.25), four fractional numbers with halves (1.50, 2.50, 3.50, 4.50), and four fractional numbers with three-fourths (1.75, 2.75, 3.75, 4.75). For each judgment, participants indicated their estimate of the rating's magnitude on the unmarked horizontal line with 1 and 5 as its ends. At the end of the study, all participants were asked standard demographic questions such as age, gender, income, and education level.

Analyses and Results

Before we report specific analyses, it might be instructive to look at the graphical depiction of means of perceived magnitudes across the three conditions. Figure 1.3A depicts the perceived magnitudes of the 17 stimuli in the three conditions. Three patterns can be observed in this figure. First, participants overestimated the magnitude of fractional star ratings and underestimated the magnitude of fractional Arabic numeral ratings. Second, participants' responses to the combination of star and Arabic numerals were almost identical to their responses to Arabic numerals only, suggesting that Arabic numerals play a dominant role when both Arabic numerals and star symbols are present. Third, the misestimation was more pronounced for fractional ratings than whole number ratings.³

To assess the statistical significance of these inferences, we calculated the deviation between perceived magnitude and actual magnitude of the ratings by subtracting the actual

³ It might be argued that the lack of misestimation for whole numbers is because we used these numbers—1, 3, and 5—in the calibration phase. However, we do not find any misestimation for the whole numbers 2 and 4, which were not included in the calibration phase. Thus, we are inclined to believe that this pattern reflects a more veridical processing of whole numbers.

Figure 1.3A: Perceived Magnitude of Ratings (Study 1)

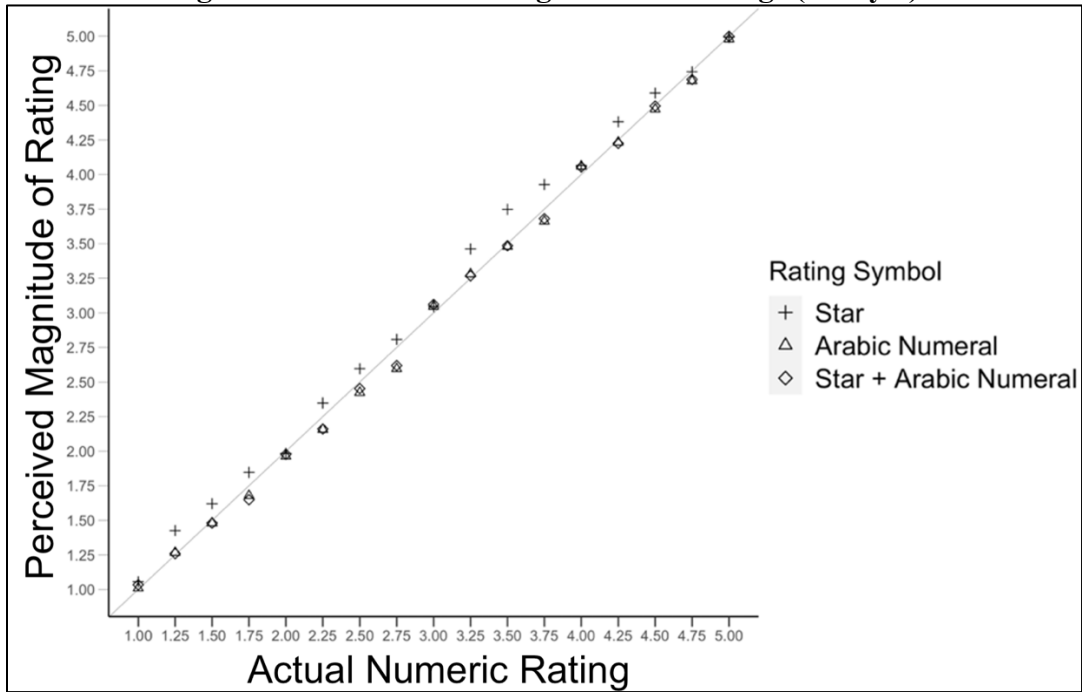
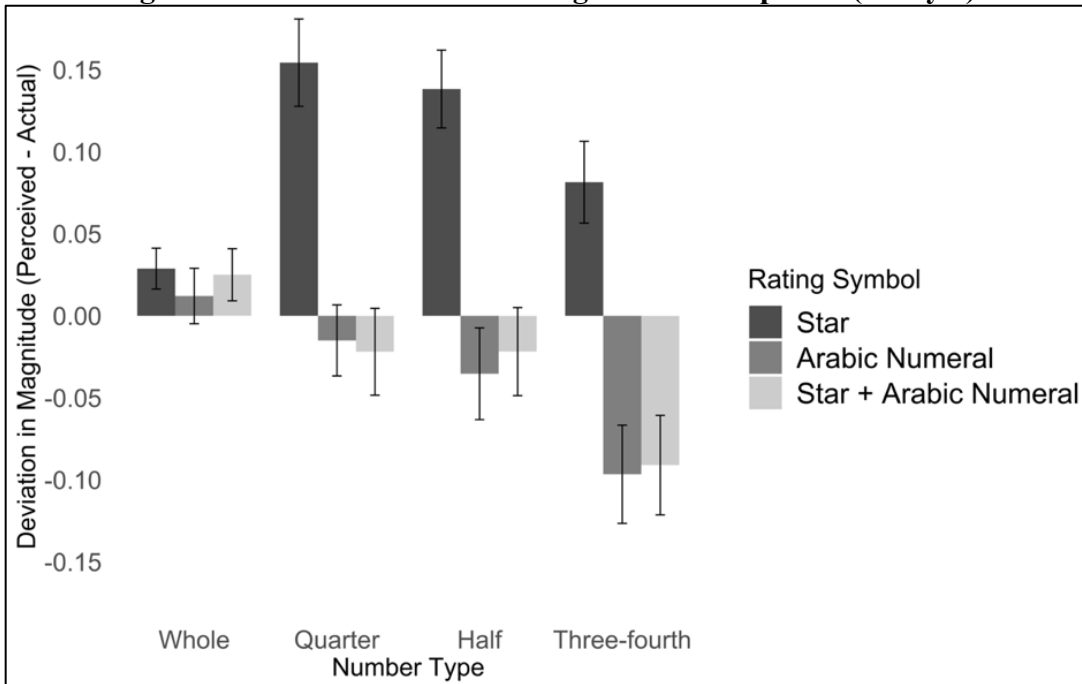


Figure 1.3B: Misestimation in Magnitude Perceptions (Study 1)



Figures 3A and 3B—These figures show the results from Study 1. Figure 1.3A shows how the average perceived magnitudes of the 17 stimuli change with different rating symbols. Figure 1.3B shows the misestimation in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) by type of number across rating symbols. The bars in Figure 1.3B represent 95% confidence intervals. Both figures depict raw means.

numerical rating from the perceived magnitude responses (i.e., perceived magnitude – actual magnitude). We averaged the deviations for whole numbers (1.00, 2.00, etc.), fractional numbers with quarters (1.25, 2.25, etc.), fractional numbers with halves (1.50, 2.50, etc.), and fractional numbers with three-fourths (1.75, 2.75, etc.) to examine their average deviations. These average deviations were analyzed using a 3 x 4 mixed factorial ANOVA with rating symbol type (star, Arabic numeral, both star and Arabic numeral) as the between-subjects factor and number type (whole, quarter, half, and three-fourths) as the within-subjects factor⁴. The main effect of rating symbol was significant ($F(2, 613) = 52.76, p < .0001, \eta^2_p = .15$), as was the main effect of number type ($F(3, 1839) = 57.51, p < .0001, \eta^2_p = .09$). Additionally, the interaction between rating symbol and number type was significant ($F(6, 1839) = 38.15, p < .0001, \eta^2_p = .11$).

Figure 1.3B displays the deviation of subjective magnitude perception from the actual magnitude of the ratings for these number types. Participants displayed a tendency to overestimate ratings when shown as fractional stars, whereas they underestimated ratings presented as fractional Arabic numerals on average⁵ ($M_{\text{Star}} = +.12 [+ .10, +.15], SD = .30$; $M_{\text{ArabicNumeral}} = -.05 [-.07, -.03], SD = .30$; $F(1, 668) = 132.64, p < .0001, \eta^2_p = .17$), providing support for H1a and H1b.

Next supporting H2, when whole numbers were used, both misestimation of stars and Arabic numerals were attenuated and no significant difference was observed in participants' estimations between star and Arabic numeral ratings ($M_{\text{Star}} = +.03 [+ .00, +.05], SD = .18$; $M_{\text{ArabicNumeral}} = +.01 [-.01, +.04], SD = .20$; $F(1, 1136) = .92, p = .34, \eta^2_p = .00$). Furthermore,

⁴ We pre-registered a 2 x 3 mixed ANOVA analysis with rating symbol (star, Arabic numeral, both star and Arabic numeral) as the between-subjects factor and number type (whole, fractional) as the within-subjects factor. All our results hold if we analyze the data using our original pre-registered ANOVA model. The reported analysis is more informative.

⁵ The underestimation for quarter Arabic numerals was not statistically significant. H4b predicts how the magnitude of underestimation varies across different fractional numbers, which accounts for this observation.

supporting H2, the overestimation of star ratings was higher for fractional star ratings ($M_{\text{FractionalStars}} = +.12$ [$+.10, +.15$], $SD = .30$) than for whole star ratings ($M_{\text{WholeStars}} = +.03$ [$+.00, +.05$], $SD = .18$; $F(1, 1839) = 119.33$, $p < .0001$, $\eta^2_p = .06$). Similarly, the underestimation of Arabic numeral ratings was greater for fractional numeric ratings; fractional Arabic numerals were underestimated ($M_{\text{FractionalArabicNumeral}} = -.05$ [$-.07, -.03$], $SD = .30$), while whole numerals were not underestimated ($M_{\text{WholeArabicNumeral}} = +.01$ [$-.01, +.04$], $SD = .20$; $F(1, 1839) = 48.85$, $p < .0001$, $\eta^2_p = .03$)⁶.

Next, we tested H4, which posited that the size of misestimation of fractional ratings will be proportional to the distance between the anchor and the fractional rating. Given that the distance between the anchor and rating is larger for quarter star ratings (e.g., 3.25) than three-fourth star ratings (e.g., 3.75), we observed a more pronounced overestimation for quarter stars ($M_{.25Stars} = +.15$ [$+.13, +.18$], $SD = .29$) compared to three-fourth stars ($M_{.75Stars} = +.08$ [$+.06, +.11$], $SD = .31$; $F(1, 1839) = 45.95$, $p < .0001$, $\eta^2_p = .02$). For Arabic numeral ratings, the pattern reversed. The underestimation was more substantial for three-fourth Arabic numerals ($M_{.75ArabicNumerals} = -.10$ [$-.12, -.07$], $SD = .32$) than for quarter Arabic numerals ($M_{.25ArabicNumerals} = -.02$ [$-.04, +.01$], $SD = .26$; $F(1, 1839) = 58.20$, $p < .0001$, $\eta^2_p = .03$). Thus, H4a and H4b are supported.

Finally, we looked at the effect of using both star ratings and Arabic numerals. There was no significant difference in participants' responses in the Arabic numeral and the combined rating conditions ($M_{\text{ArabicNumeral}} = -.03$ [$-.05, -.01$], $SD = .27$; $M_{\text{Star+ArabicNumeral}} = -.03$ [$-.05, -.01$],

⁶ In our studies, we observe a consistent pattern where participants slightly overestimate whole numbers, regardless of the rating symbols used. This phenomenon aligns with prior research findings that suggest round numbers are often perceived as having a greater magnitude than fractional numbers in various contexts. For instance, Thomas, Simon, and Kadiyali (2010) discovered that precise house prices (e.g., \$395,425) are perceived as lower than rounded prices of similar value (e.g., \$395,000). Similarly, Pena-Marín and Bhargava (2016) observed that consumers perceive a caffeine drink containing 200 mg (as opposed to 203 mg) of caffeine as providing more prolonged energy.

SD = .27; $F(1, 613) = .18, p = .67, \eta^2_p = .00$); see Figure 1.3B.

Response times. We also evaluated participants' response times for each magnitude judgment task. We averaged the response times for each number type and then analyzed the logarithmically transformed average response times using the same 3 x 4 mixed factorial ANOVA as in the magnitude judgment analysis. The results revealed that the type of rating symbol used—be it star symbols, Arabic numerals, or a combination thereof—did not significantly influence the response times ($F(2, 613) = .86, p = .43, \eta^2_p = .00$). Additionally, no interaction was observed between the rating symbol and the number type ($F(6, 1839) = 1.11, p = .35, \eta^2_p = .00$). However, a significant effect was found regarding the number type ($F(3, 1839) = 97.24, p < .0001, \eta^2_p = .14$). Specifically, the response times for fractional values ($M_{\log RT_{.25}} = 1.24, SD = .72$; $M_{\log RT_{.50}} = 1.26, SD = .73$; $M_{\log RT_{.75}} = 1.25, SD = .74$) were significantly higher in comparison to whole numbers ($M_{\log RT_{\text{Whole}}} = 1.03, SD = .67$; $F(1, 1839) = 176.33, p < .0001, \eta^2_p = .14$). This finding supports our theory that fractional numbers are relatively harder to process as opposed to whole numbers.

Discussion

Several important results emerge from Study 1. First, it was observed that individuals generally overestimate fractional numerical quantities when represented by star symbols and underestimate them on average when presented as Arabic numerals, thereby supporting H1a and H1b. Next, in line with H2, this tendency to misestimate was reduced for whole numbers, resulting in no significant differences in magnitude estimations across the different rating symbol types for whole numbers.

The pattern of results also offers insights into the mechanisms underlying such misestimations. Supporting H4a and H4b, we found that the overestimation in star symbols was stronger for quarter ratings than for three-fourths ratings. Conversely, the underestimation in Arabic numerals was stronger for three-fourths ratings than for quarter ratings. This shows that the further the anchor is from the actual number, the greater the deviation from the actual rating is. Thus, these results support our rating symbol contingent anchoring hypothesis.

Furthermore, Study 1 revealed that when participants made magnitude judgments using both star symbols and Arabic numerals, their evaluations aligned closely with those using only Arabic numerals. This finding may highlight the influential role of Arabic numerals in the encoding process. It appears that in contexts where both star ratings and Arabic numerals are presented, the latter may inadvertently dominate the judgment process. This phenomenon likely stems from the ubiquitous use of Arabic numerals in denoting numerical quantities, which renders them more familiar and thus more readily processed by the brain for making numerical judgments.

Importantly, this study also rules out the alternative hypothesis that participants rounded fractional star ratings to the nearest whole number due to cognitive shortcuts. If rounding was the explanation, we would expect an underestimation of quarter ratings due to rounding down and an overestimation of three-fourths ratings because of rounding up. However, our results showed overestimation of both quarter and three-fourth star ratings, consistent with our theory of contingent anchoring. These findings provide further evidence for the influence of initial anchors on subsequent judgments, rather than simple rounding, as an explanation for the observed effects.

Supplementary Studies. In Study 1 we used stars to represent graphical symbols, a rating symbol that is popular across various platforms. Nevertheless, a question arises whether the observed overestimation is applicable to other graphical representations. Perhaps stars have a positive halo, which might be influencing people’s magnitude judgments. However, our theory suggests that this overestimation is attributable to visual completion of partially complete graphic symbols, implying that even incomplete circles or squares would be similarly overestimated. To explore this, we replicated Study 1 using circles in two follow-up studies that used circles and squares. The details of these two studies are available as Supplementary Study 1 and Supplementary Study 2 respectively in the Chapter 1 Appendix.

Figure 1.4: Misestimations in Magnitude Perceptions Using Circles as Rating Symbols

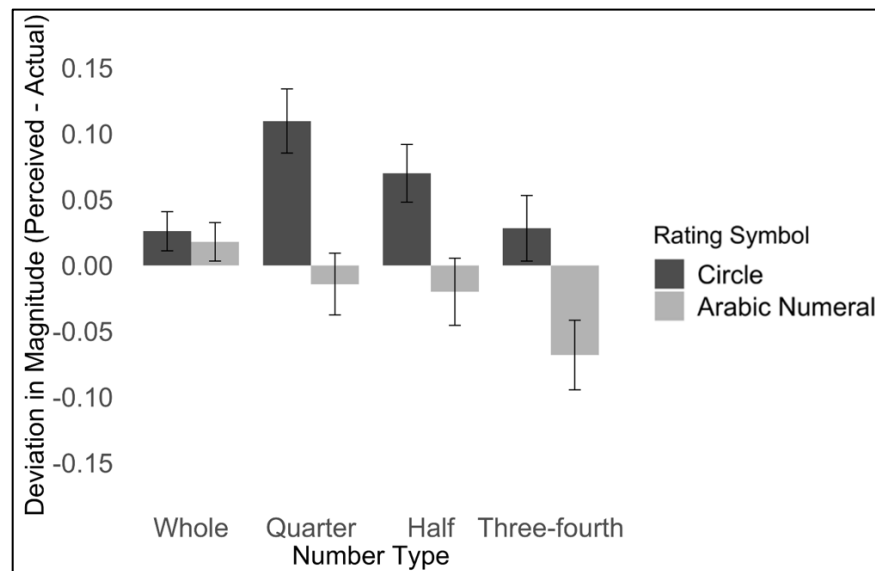


Figure 1.4—This figure shows the misestimation in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) by type of number across rating symbols. The bars represent 95% confidence intervals, and the plot depicts raw means.

Figure 1.4 summarizes the results from Supplementary Study 1 where we used circles instead of stars. As depicted, the results from this study using circles replicate those from Study 1 using stars. Similarly, Supplementary Study 2 where we used squares as graphical symbols, once again replicated these results, confirming a consistent trend of overestimation in magnitude

judgments across various graphical representations, including stars, circles, and squares.

A notable limitation of our initial studies is the lack of direct evidence on whether individuals use rounded-up numbers as anchors for graphical ratings and rounded-down numbers as anchors for Arabic numeral ratings. We designed the subsequent study to directly address this question.

STUDY 2: DIFFERENT ANCHORS

In Study 2, we aimed to verify a critical assumption in our theory, namely, that people tend to instinctively round up fractional star ratings and round down fractional Arabic numeral ratings, creating anchor points that influence their subsequent judgments. To test this, we analyzed participants' rounding tendencies for half ratings presented either as star symbols or Arabic numerals. This study was pre-registered at OSF (<http://bit.ly/3PASudi>).

Participants and Procedure

In this study, we utilized 321 participants (38.2% female, 59.6% male, 2.1% nonbinary, $M_{\text{age}} = 40$ years) from the CloudResearch panel, composed of U.S. residents, who received monetary compensation for their participation. Adhering to our pre-registered exclusion criteria, we excluded choice responses where neither the rounded-up nor rounded-down rating values were selected ($n = 7$), resulting in an analysis of 1277 choice responses. For detailed information on data exclusions, please refer to the Chapter 1 Appendix.

Participants were randomly assigned to either the star or Arabic numeral rating symbol condition. In each condition, they were shown four half ratings (1.5, 2.5, 3.5, 4.5) in a random order, one at a time, and asked to indicate how they would describe this rating to others. Those in

the star rating condition saw these ratings depicted using images of stars (e.g., an image of three and a half stars). Those in the Arabic numeral condition saw the ratings in Arabic numerals (e.g., “3.5 stars”). Participants responded by picking one of the five options: “around 1 star”, “around 2 stars”, “around 3 stars”, “around 4 stars”, “around 5 stars” (see Chapter 1 Appendix for details). They were informed that there were no right or wrong answers and that they should answer based on their instincts. Note that these response options required participants to round up or down the fractional ratings. We chose ratings ending in halves (e.g., 3.5) as stimuli because they provided equally valid options for rounding up or down, allowing us to test how the rating symbols influence participants’ initial instincts in this regard.

All participants were asked standard demographic questions such as age, gender, income, and education level at the end of the study.

Table 1.1: % Choice of Rounded-up (vs. Rounded-down) Anchor in Study 2

Numeric Rating	Star Symbols	Arabic Numeral Symbols
1.5	82%	29%
2.5	80%	29%
3.5	78%	29%
4.5	75%	30%
Average	79%	29%
95% CI	[72%, 85%]	[22%, 37%]

Table 1.1—This table shows the summary of results for Study 2. The values are estimated from models reported in the text.

Analyses and Results

To test our theory, we coded participants’ responses as a binary variable (1 = rounded-up, 0 = rounded-down) to record whether they rounded-up or rounded-down the fractional ratings. For example, if a participant described 3.5 as “around 4 stars,” it was coded as 1, and if they described it as “around 3 stars,” it was coded as 0. We then submitted this binary measure to a mixed-effects logistic regression that accounted for the random effect of participants and

included rating symbol (star vs. Arabic numeral; coded as a dummy variable) and numeric rating (1.5, 2.5, 3.5, 4.5; standardized with mean zero and standard deviation of one) as predictor variables, along with their two-way interaction term.

Consistent with our prediction, the effect of rating symbol was significant ($\beta = 2.21$, $SE = .29$, $z = 7.70$, $p < .0001$, $d = 1.22$). Participants rounded up 79% of the half star ratings ($M_{\text{Star}} = .79$ [.72, .85], $SE = .03$). However, only 29% of the same half ratings were rounded up when expressed in Arabic numerals ($M_{\text{ArabicNumeral}} = .29$ [.22, .37], $SE = .04$). The two-way interaction ($\beta = .17$, $SE = .15$, $z = 1.14$, $p = .25$, $d = .09$) was not significant, indicating that the effect was consistent across all rating values (see Table 1.1).

Discussion

Study 2 shows that people tend to round up fractional star ratings, whereas round down identical fractional Arabic numerals. This pattern of results is consistent with our theory: magnitude judgments are anchored on the rounded-up numbers when encoding fractional graphic ratings but are anchored on rounded-down numbers for the same fractional values expressed in Arabic numerals.

STUDY 3: ROLE OF VISUAL INCOMPLETENESS

Why do people overestimate fractional stars? This study was designed to address this question. Our theory posits that the tendency to visually complete incomplete stimuli is the root cause of the overestimation of fractional graphical ratings. This tendency is activated by visual cues of incompleteness. Previous studies have demonstrated that providing visual cues of completion can alleviate this tendency (Gerbino 2020; Kanizsa 1979), although it might not

completely eliminate it. For example, when viewing a picture of a human body with a missing hand, the visual completion process is activated, which makes people incorrectly recall seeing the complete hand in the picture. But this propensity to incorrectly complete the picture of human body is reduced when there is evidence of amputation, which makes the picture visually complete (Kanizsa 1979). Based on this prior work, we hypothesized that visual cues of completion might reduce the overestimation of fractional star ratings (H3). As noted previously, since visual completion is an automatic process, we did not expect this intervention to completely remove the overestimation; we expected it to reduce the extent of overestimation.

We used two different types of star ratings in this study, *visually complete* stars and *visually incomplete* stars. Using the same number line estimation task as in the first two studies, we compared the magnitude judgments for visually incomplete and visually complete star symbols. This study was pre-registered at OSF (<http://bit.ly/3uZUOBd>).

Participants and Procedure

We recruited participants using a US standard sample from Prolific in exchange for monetary compensation and used the data from 543 participants (54.6% female, 44.6% male, .7% nonbinary, $M_{\text{age}} = 42$ years). See Chapter 1 Appendix for details on data exclusions.

This study employed a mixed design where type of stars (visually complete vs. visually incomplete) was a between-subjects factor and rating symbols (star vs. Arabic numeral) was a within-subjects factor. Participants were randomly assigned to one of two groups. One group was asked to estimate the magnitude of 17 visually complete stars, while the other group was asked to estimate the magnitude of 17 visually incomplete stars (see Study 3 stimuli in Figure 1.2) on

an unmarked number line. We used the same paradigm as in the first study, i.e., calibration phase followed by the mental number line task for 17 numerical quantities.

In addition to its primary aim, this study had a secondary objective: testing the robustness of the divergence between star ratings and Arabic numerals using a within-subjects design.

Participants in this study had to respond to two sets of ratings within a span of minutes.

Immediately after they evaluated the star ratings, we asked participants to evaluate the same ratings in Arabic numerals. Using this within-subjects approach, we examined whether the overestimation of star ratings in the first task had an impact on their subsequent judgments of Arabic numeral ratings. We hypothesized that these misestimations, like perceptual illusions, are largely driven by salient perceptual representations. Therefore, even after estimating star ratings, and regardless of whether the stars are visually complete or incomplete, participants would underestimate the Arabic numeral ratings.

Thus, in summary, the same 17 ratings were first presented as complete or incomplete star ratings. One group evaluated visually complete stars, while the other group evaluated visually incomplete stars. Immediately after that both groups evaluated the same 17 ratings represented as Arabic numerals.

At the end of the study, all participants answered standard demographic questions such as age, gender, income, and education level.

Analyses and Results

We calculated the deviation between the perceived magnitude of the ratings and their actual magnitude. We averaged the deviations for whole numbers, fractional numbers with quarters, fractional numbers with halves, and fractional numbers with three-fourths and submitted the average deviations to a 2 x 4 x 2 mixed ANOVA with type of stars (visually

complete vs. visually incomplete) as the between-subjects factor and number type (whole, quarter, half, and three-fourth) and rating symbol type (star vs. Arabic numeral) as within-subjects factors. All three main effects were significant (type of stars: $F(1, 541) = 16.57, p = .0001, \eta^2_p = .03$; number type: $F(3, 1623) = 121.81, p < .0001, \eta^2_p = .18$; and rating symbol type: $F(1, 541) = 163.18, p < .0001, \eta^2_p = .23$), along with a significant three-way interaction ($F(3, 1623) = 21.06, p < .0001, \eta^2_p = .04$).

The left panel of Figure 1.5 shows how providing visually complete stars influenced participants' magnitude perceptions. Our interest is in comparing the magnitude perceptions of visually complete and incomplete stars for each type of number⁷. For quarter stars, using visually complete stars reduced overestimation from $+0.15 [+0.13, +0.17]$ to $+0.10 [+0.08, +0.12]$ ($F(1, 1760) = 10.12, p = .002, \eta^2_p = .006$). This suggests that, compared to visually incomplete quarter stars, fewer participants rounded up the visually complete quarter stars to full stars. Alternatively, these results also suggest the possibility that participants used half stars as anchors to judge the magnitude of visually complete quarter stars.

In the case of half stars, using visually complete stars completely reduced the overestimation from $+0.12 [+0.10, +0.14]$ to $+0.01 [-0.01, +0.03]$ ($F(1, 1760) = 47.46, p < .0001, \eta^2_p = .03$). This result suggests that while visually incomplete stars were rounded up to full stars by participants, this rounding was not observed with visually complete stars. Instead, participants appeared to anchor their judgments on the half star, leading to remarkably precise estimations. That is, when they saw three and half visually complete stars, they used 3.5 as their initial anchor.

⁷ Averaging across all fractional numbers (quarters, halves, and three-fourths), using visually complete stars completely eliminated the overestimation from $+0.11 [+0.10, +0.13]$ to $+0.015 [-0.00, +0.03]$ ($F(1, 958) = 59.88, p < .0001, \eta^2_p = .06$).

For three-fourth stars, the use of visually complete stars shifted the misestimation in magnitude from an overestimation of $+0.08$ [$+0.06$, $+0.10$] to an underestimation of -0.07 [-0.09 , -0.05] ($F(1, 1760) = 91.97, p < .0001, \eta^2_p = .05$). We suspect that for visually complete three-fourth stars, many participants might have used the nearest half stars as anchors. That is, a visually complete star depicting 3.75 (three complete stars and a fourth star that is 75% of a full star), might have triggered an anchor of 3.5, resulting in an underestimation of its magnitude. Note that although we did not predict this result a priori, it is completely consistent with our hypothesis that people are less likely to fill up visually complete stars and use rounded-up numbers as initial anchors.

Figure 1.5: Effect of Visual Completeness by Number Type (Study 3)

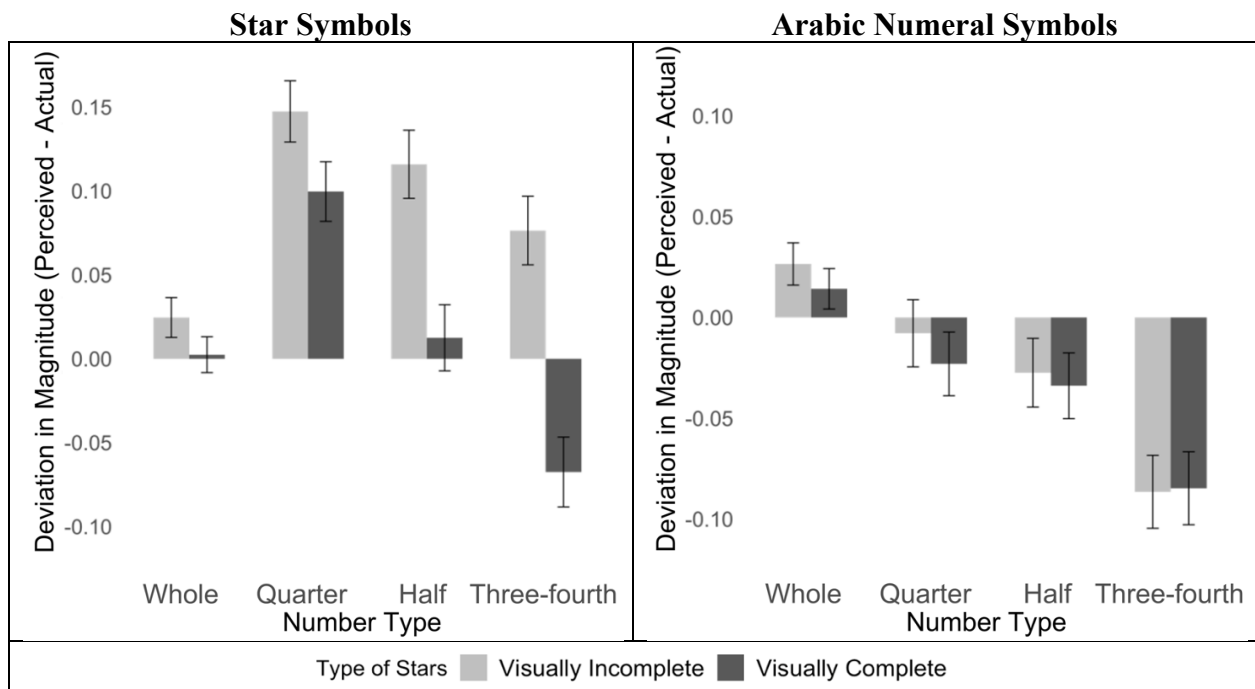


Figure 1.5—This figure shows the results of Study 3 by the four number types: quarters, halves, three-fourths, and whole numbers. The plot shows the misestimation in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) of star ratings (left panel) and Arabic numeral ratings (right panel) by the type of stars used. The bars represent 95% confidence intervals. The plot depicts raw means.

Now let us turn our attention to the right panel, depicting the magnitude perception of

Arabic numerals. Recall that all participants responded to the Arabic numeral magnitude judgment task immediately after evaluating either visually complete stars or incomplete stars. Our intention here was to see whether their perceptions of star ratings would in any way carry over and influence their perception of Arabic numerals. We found that fractional Arabic numeral ratings were on average underestimated, irrespective of whether the stars were visually incomplete ($M_{\text{VisuallyIncomplete}} = -.04 [-.06, -.02]$, $SD = .29$) or complete ($M_{\text{VisuallyComplete}} = -.05 [-.06, -.03]$, $SD = .28$; $F(1, 958) = .27$, $p = .60$, $\eta^2_p = .00$). This suggests that the effect of rating format on magnitude judgments are largely stimulus-driven with little carryover effects. When participants see a fractional star rating, they might or might not overestimate it depending on whether the star is complete or incomplete, but when they see a fractional Arabic numeral rating, they tend to underestimate it regardless of their response to the star rating.

Discussion

Replicating the pattern observed in Study 1, in this study we find that participants overestimate ratings when represented by visually incomplete stars. Notably, this trend of overestimation was significantly reduced when stars were presented as visually complete, thereby lending support to H3. Additionally, Study 3 demonstrates that even after exposure to star ratings, the tendency to underestimate fractional Arabic numeral ratings persists and is robust. The misestimations of ratings caused by rating formats seem to be largely a context effect driven by the stimulus. Even when stars and Arabic numerals are interspersed, consumers might overestimate stars and underestimate Arabic numerals.

The initial three studies collectively unveil two distinct perceptual phenomena influencing subjective magnitude perceptions of numeric ratings: the visual completion effect and the left-digit effect. However, we haven't investigated if these phenomena could also impact consumers' recall of the actual value of the numeric rating. We designed Study 4 to test this.

STUDY 4: MEMORY RECALL OF RATINGS

How does the rating format influence the recalled magnitudes of the ratings? Addressing this question has both practical and theoretical implications, providing us more insight into how different symbolic representations influence magnitude judgments.

In everyday life, people usually use the Arabic numeral format to think of magnitudes. When consumers remember or talk about magnitudes, they typically do so in Arabic numerals. Our conceptualization predicts that when consumers see star ratings and later try to recall the star ratings in Arabic numerals, their recall will be overestimated. When participants are first shown star ratings and later asked to recall them as numerals, the process involves two steps: first, the star ratings have to be mapped onto an analog magnitude, and then this magnitude has to be converted into an Arabic numeral. Because the analog magnitude representations of star ratings will be overestimated, we expect that the recalled values of star ratings will also be overestimated.

In contrast, when the stimulus format and the recall format are identical, we do not expect any misestimation. That is, when the ratings are presented as Arabic numerals and recalled as Arabic numerals, their recall is less likely to be misestimated. If ratings are presented and recalled as Arabic numerals, the recall process is straightforward without any scope for perceptual distortions.

Study 4 was designed to test this hypothesis. Participants were shown products with various ratings and were later asked to recall these ratings using an Arabic numeral slider scale. Half of the participants viewed star ratings, and the other half saw numeral ratings. The objective was to assess whether the initial format of the ratings influenced recall accuracy. The methodology for this study was pre-registered at OSF (<http://bit.ly/3V7KXnz>).

Participants and Procedure

We employed 355 participants (52.5% female, 47% male, .5% nonbinary, $M_{\text{age}} = 39.6$ years) from the CloudResearch panel of US participants in exchange for monetary compensation. See Chapter 1 Appendix for details on data exclusions based on pre-registered criteria. Participants were shown four perfumes listed on an e-commerce website, along with their average ratings out of five. Participants were randomly assigned to either the star or Arabic numeral rating symbol condition (see Study 4 stimuli in Figure 1.2) and the symbols used to describe the ratings were according to the assigned condition. Participants were shown four perfumes, each with a different rating (e.g., 1.5, 2.5, 3.5, 4.5) in a random order. After seeing each perfume, participants were asked to recall the name and average rating of each perfume on the subsequent screen. Participants used a text box to recall the name of the perfume and used a numbered slider scale with endpoints 0 and 6 to recall the rating of the perfume. The slider scale allowed participants to see the exact value of their response (see Chapter 1 Appendix for a visual illustration).

At the end of the study, we verified that participants were aware that the product ratings shown were out of five and asked participants standard demographic questions such as age, gender, income, and education level.

Analyses and Results

Deviation in Recall Magnitudes. To analyze the errors in recall for graphical and Arabic numeral ratings, we calculated the deviation between the recalled and actual magnitude of the rating by subtracting the actual numerical rating from the recalled response (i.e., recalled magnitude minus actual magnitude). We then subjected the deviations to a 2 x 4 mixed-factorial ANOVA, with rating symbol (star vs. Arabic numeral) as the between-subjects factor and rating value (1.5, 2.5, 3.5, 4.5) as the within-subjects factor. Only the main effect of the rating symbol was significant ($F(1, 347) = 20.40, p < .0001, \eta^2_p = .06$). The interaction was not significant ($F(3, 1044) = .79, p = .50, \eta^2_p = .00$). The analysis results showed a significant overestimation in participants' recalls of star ratings, with a mean of $+0.14$ [$+0.10, +0.18$] and a standard deviation of $.56$. However, there was no average significant error in the recalls of Arabic numeral ratings ($M_{\text{ArabicNumeral}} = +0.02$ [$-0.02, +0.06$], $SD = .37$). See Table 1.2 for detailed results.

Table 1.2: Average Recalled Ratings in Study 4

Numeric Rating	Star Symbols	Arabic Numeral Symbols
1.5	1.72 {22%, 32%}	1.59 {9%, 10%}
2.5	2.70 {20%, 34%}	2.53 {12%, 12%}
3.5	3.62 {11%, 44%}	3.52 {4%, 15%}
4.5	4.52 {14%, 41%}	4.45 {9%, 12%}
Average (3.0)	3.14 {17%, 38%}	3.01 {9%, 12%}
95% CI	[3.10, 3.17]	[2.98, 3.05]

Table 1.2—This table presents a summarized overview of findings from Study 4, detailing the average recalled ratings and the proportion of participants whose estimates fell below or above the actual values. Enclosed within curly braces, the data is organized as follows: the first number denotes the percentage with deviations less than -0.01 (underestimates), and the second denotes the percentage with deviations greater than $+0.01$ (overestimates).

Discussion

Recalls of star ratings tended to exceed their true values, whereas those of Arabic numerals were accurate on average. This discrepancy has implications for consumer behavior,

particularly in how product ratings are remembered and communicated through word-of-mouth. The choice between star and Arabic numeral ratings by retailers may significantly influence the consumer's recall accuracy, potentially shaping their purchase decisions and the recommendations they share with others based on those recalls.

While the first four studies have shed light on the cognitive processes influencing misestimations in magnitude judgments of numerical quantities, the broader impact of these misestimations on consumers' judgments and decisions remains an area for further exploration. The next three studies are thus designed to delve into the downstream consequences of these misestimations.

STUDY 5: WILLINGNESS TO PAY

How does the format of the rating influence consumers' valuation of products? In Study 5, we investigated the impact of different rating symbols on participants' quality judgments and their willingness to pay (WTP) for products. Participants were shown several ballpoint pens and their average ratings, which were either presented as star symbols or Arabic numerals. They were then asked to indicate their WTP for the pens and provide their quality judgments of the pens. Study 5 was pre-registered at OSF (<http://bit.ly/3BJdc5a>).

Participants and Procedure

We used 334 participants (46.9% female, 52.3% male, .9% nonbinary, $M_{\text{age}} = 41$ years) from the CloudResearch panel of US participants in exchange for monetary compensation. See Chapter 1 Appendix for details on data exclusions based on pre-registered criterion. We randomly assigned participants to either the star rating or Arabic numeral rating condition (see Study 5 stimuli in Figure 1.2). Participants were shown several ballpoint pens listed on an e-

commerce website, along with their average ratings out of five (see Chapter 1 Appendix for stimuli details). The symbols used to describe the ratings depended on the rating condition assigned. In each condition, we presented four ballpoint pens, each with a different rating (i.e., 1.5, 2.5, 3.5, 4.5), in a random order, one at a time. For each pen, we asked participants to indicate their WTP for the pen (between \$1 and \$10) using a text box. After indicating their WTP for the four pens, participants then rated the quality of each pen using a four-point scale (1 = “very low,” 2 = “low,” 3 = “high,” and 4 = “very high”).

At the end of the study, we verified that participants were aware that the product ratings shown were out of five and asked participants standard demographic questions such as age, gender, income, and education level.

Table 1.3: Willingness to Pay and Quality Judgments in Study 5

Numeric Rating	Willingness to Pay (\$)		Quality Judgments	
	Star Symbols	Arabic Numeral Symbols	Star Symbols	Arabic Numeral Symbols
1.5	\$1.41	\$1.28	1.32	1.30
2.5	\$1.95	\$1.61	2.04	1.98
3.5	\$3.07	\$2.53	2.91	2.74
4.5	\$4.09	\$3.96	3.65	3.52
Average	\$2.63	\$2.35	2.48	2.38
95% CI	[\$2.48, \$2.78]	[\$2.20, \$2.50]	[2.43, 2.52]	[2.34, 2.43]

Table 1.3—This table shows summary of results for Study 5. The values are estimated from models reported in the text.

Analyses and Results

To investigate the impact of rating symbols on participants’ willingness to pay and quality evaluations, we conducted two 2 x 4 mixed factorial ANOVAs. In each ANOVA, we included rating symbol (star vs. Arabic numeral) as the between-subjects factor and rating value (1.5, 2.5, 3.5, 4.5) as the within-subjects factor. The results showed that participants were willing to pay, on average 12% more, more for pens with star ratings compared to those with Arabic

numeral ratings ($M_{\text{Star}} = \$2.63$, $SD = 1.77$; $M_{\text{ArabicNumeral}} = \$2.35$, $SD = 1.57$; $F(1, 325) = 6.85$, $p = .009$, $\eta^2_p = .02$). Additionally, participants rated pens with star ratings to be of higher quality than those with Arabic numeral ratings ($M_{\text{Star}} = 2.48$, $SD = 1.01$; $M_{\text{ArabicNumeral}} = 2.38$, $SD = .98$; $F(1, 332) = 7.60$, $p = .006$, $\eta^2_p = .02$). See Table 1.3 for details.

Mediation Analysis. We aggregated individual participants' WTP and quality judgment ratings by calculating the mean across the four responses provided by each participant. We then implemented a mediation analysis (Hayes, 2017, Model 4) with the rating symbol serving as the independent variable, the averaged WTP as the dependent variable, and the mean quality evaluation as the mediating variable. Quality evaluation fully mediated the relationship between the rating symbol and WTP (standardized indirect effect: $B = .11$; 95% CI = [.05, .18]). See Chapter 1 Appendix for a visual representation of the mediation model.

Discussion

Study 5 demonstrates that the representation format of product ratings—graphical symbols versus Arabic numerals—significantly impacts consumer judgment and purchase decisions. Graphical symbols, as compared to Arabic numerals, lead to an overestimation of rating magnitude, influencing product quality inferences and, consequently, consumers' willingness to pay higher prices.

STUDIES 6A AND 6B: SOCIAL MEDIA AD ENGAGEMENT

Can rating formats influence engagement with social media campaigns? Studies 6A and 6B examine how rating symbols influence consumer engagement with social media product advertisements. To this end, we utilized the research capabilities of Meta platforms, including

Facebook and Instagram, to conduct our field experiments. These Meta field experiments⁸ harness the algorithms and extensive user base of these platforms to measure real-world responses to controlled variations in advertising content—an approach that has been employed in prior research (Rifkin, Du, and Cutright 2023; To and Patrick 2021).

The advertising campaign featured a fictitious succulent plant company named Lushleaf, depicted in Figure 1.6. This campaign included three distinct ad conditions, differing solely in the representation of rating symbols: one using star symbols, one using the short version of Arabic numerals, and a third one using the longer version of Arabic numerals. We compared ad engagement, quantified through the number of ad clicks, across the ad conditions.

Figure 1.6: Ad Campaigns for A/B/C Tests in Study 6A



In Study 6A, we used a product with a rating of 3.5. To validate the consistency of our results across different rating magnitudes, Study 6B replicated the exact methodology of Study 6A but with a product rating of 2.5. Both studies were pre-registered at OSF (<https://bit.ly/3MNqCIW> and <https://bit.ly/47IQgk5>).

⁸ More information on meta experiments is provided here- <https://www.facebook.com/business/help/3506622486044209>. It is important to acknowledge that Meta A/B test methodology does not ensure ‘true’ randomization, as platform algorithms optimize ad visibility post-randomization, which potentially limits the ability to draw causal inferences (Boegershausen et al., 2023). To address this concern, and ensure the robustness of our results, we ran the study twice, using slightly different stimuli.

Participants and Procedure

Meta users based in the United States were randomly allocated to one of three groups in a between-subjects design (rating symbol type: star, numeral short, or numeral expanded). The short numeral condition displayed the rating magnitude alongside a single star symbol, while the expanded numeral expressed ‘star’ in text and indicated that the rating was out of 5; see Figure 1.6. Utilizing the A/B test functionality of Meta Ads Manager, which enables marketers to compare multiple messages or "creatives" while controlling for other variables, we conducted two field experiments. Study 6A was carried out over a period of seven days, from May 14 to May 20, 2023, and Study 6B was conducted across two days, from November 9 to November 10, 2023. These studies collectively involved 44,994 unique Meta account users (22,220 in Study 6A and 22,774 in Study 6B). Both studies had identical experimental settings, including an A/B test on creative; objective: traffic, age: 18–65; location: United States; language: English; all devices; an optimization goal centered on link clicks; a bid strategy aimed at achieving the highest volume; an auction-based buying type; no limitations regarding Wi-Fi connectivity; and ad placements featured in Feeds, Stories, and Reels.

The advertising campaign featured a fictitious succulent plant company named Lushleaf. This campaign included three ad conditions, distinguished only by rating symbols: star, numeral short, and numeral expanded. Using two forms of Arabic numeral symbols served a secondary purpose in this study. Ratings displayed using graphical symbols include the information of the maximum rating possible (i.e., five stars in this case) while the Arabic numeral ratings (e.g., 3.5) do not include this information. It is possible that the effect of rating symbol on people’s rating evaluations is at least partially driven by this difference. This study design allowed us to test this possibility.

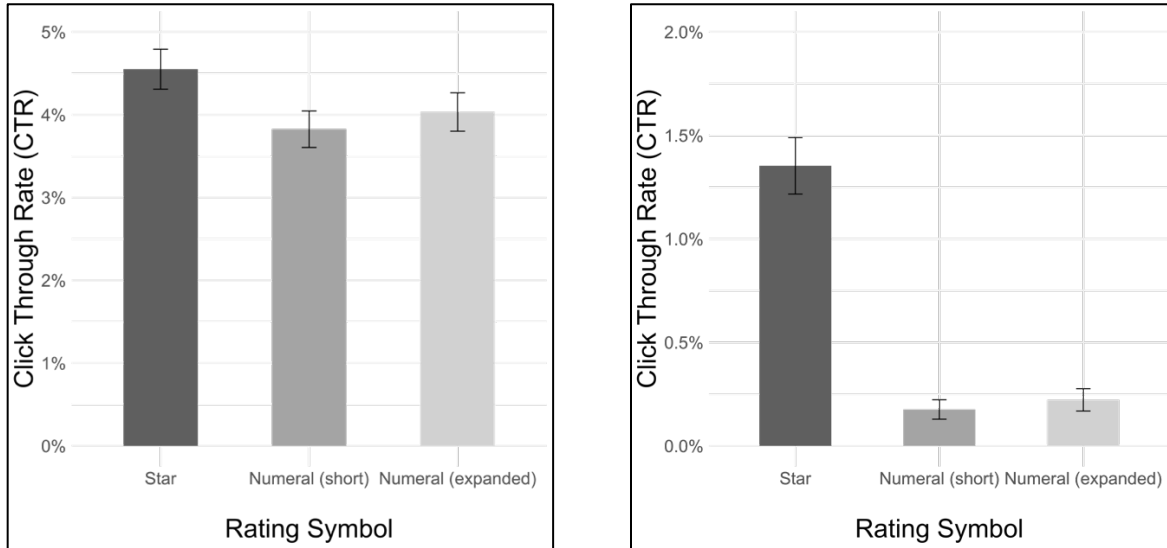
We measured engagement through the click-through rates (CTR) of these ads. Users who clicked on the ad were redirected to an Etsy search results page offering succulents from other brands (link: https://www.etsy.com/search?q=succulents&ref=auto-1&as_prefix=succ). We predicted that the ad with the star-symbol rating would generate a higher CTR than those with Arabic numeral ratings, attributed to a subjective overestimation of the rating's magnitude and a consequent perception of higher product quality.

Analyses and Results

The click-through rate (CTR) is calculated as the number of ad clicks relative to the ad's reach, which is the total number of unique Accounts Centre accounts that viewed the ad at least once. To analyze the influence of rating symbols on CTR, we created a dataset with the number of rows equal to the total reach across the three rating symbol conditions. This dataset included two columns: one specifying the rating symbol condition and the other a binary variable indicating click activity, '1' denoting a click and '0' for no click.

The star symbol condition demonstrated the highest click propensity, achieving a CTR of 339 out of 7,450 reached accounts, or 4.55%. The numeral expanded condition recorded a CTR of 290 out of 7,190, which is 4.03%, while the numeral short condition had the lowest CTR at 290 out of 7,580, or 3.82%. A logistic regression revealed a significant effect of the rating symbol on the likelihood of a user clicking on the ad. There was no significant difference in click propensity between the numeral expanded and numeral short conditions ($\beta = .055$; $SE = .08$; $z = .65$; $p = .52$; $d = .03$). However, the star symbol condition significantly outperformed the two numeral conditions in click propensity ($\beta = .15$; $SE = .07$; $z = 2.20$; $p = .028$; $d = .08$), as depicted in Figure 1.7.

Figure 1.7: Rating Symbols Influence CTR of Digital Ads (Studies 6A and 6B)



Study 6A, Rating = 3.5

Study 6B, Rating = 2.5

Figure 1.7— This figure displays the click-through rates (CTR) for Studies 6A and 6B across three ad campaigns, calculated as the ratio of clicks to unique Accounts Centre accounts that viewed the ad at least once, with bars indicating standard errors.

Study 6B replicated the findings of Study 6A, albeit with lower overall CTRs, due to the diminished appeal of the product associated with a lower rating (recall that the product used in Study 6B had a 2.5 rating). The star symbol condition again led in terms of click propensity, with a CTR of 97 out of 7,168, or 1.35%. The numeral expanded condition followed with a CTR of 17 out of 7,648, or 0.22%, and the numeral short condition had the least with a CTR of 14 out of 7,958, or 0.18%. The difference in click propensity between the two numeral conditions was not significant ($\beta = .23$; $SE = .36$; $z = .65$; $p = .52$, $d = .13$). Conversely, the click propensity in the star symbol condition was significantly greater than that in the two numeral conditions ($\beta = 1.93$; $SE = .21$; $z = 9.32$; $p < .0001$, $d = 1.07$), confirming the pattern observed in Study 6A. Additionally, note that the results of Studies 6A and 6B are robust when impressions are used in place of reach to calculate CTR.

Discussion

Studies 6A and 6B extend Study 5's findings, underscoring the real-world impact of numerical symbols on consumer behavior. These findings suggest the presentation format of ratings significantly influences user engagement. By testing the effects in a natural setting, these studies not only offer ecological validity but also quantify the impact of rating symbols on social media ad engagement.

Moreover, our findings indicate that the inclusion of maximum possible rating information does not significantly influence ad engagement. This likely stems from the widespread understanding that ratings, particularly in digital contexts, are out of 5. Therefore, explicitly stating this information seems to have no substantial impact on how consumers evaluate ratings.

GENERAL DISCUSSION

Although product ratings have become ubiquitous in online platforms, somewhat surprisingly, there has been no attempt to standardize the ratings. The format of the ratings varies across platforms. Some platforms use graphic symbols, like stars or circles, to represent ratings, while others use Arabic numerals. And some use a combination of both. Despite the wide prevalence of numerical ratings and their influence on consumer behavior, the extant research does not address the effects of using stars versus Arabic numeral ratings on consumer behavior. The current research is one of the first to examine the differential effects of graphical and Arabic numeral ratings on consumers' evaluative judgments.

Our studies uncover a reliable pattern of misestimation in people's perception of numeric ratings, particularly fractional ones. Fractional ratings, when presented graphically (such as an

image depicting three and a half stars), are typically overestimated. In contrast, the same ratings when presented in Arabic numerals (like 3.5) tend to be underestimated. Moreover, this overestimation of graphical ratings is more pronounced for quarter ratings (e.g., an image showing three and a quarter star) than for three-fourth ratings (e.g., an image showing three and three-fourth stars). Conversely, the underestimation is more substantial for three-fourth ratings in Arabic numerals than for quarter ratings. These findings lend support to our contingent anchoring framework, which posits that contingent on the type of symbols used, specific mechanisms drive these misestimations in magnitude judgments of fractional ratings. When graphic symbols such as star ratings are used, the magnitude judgment is influenced by visual completion that prompts consumers to use complete stars as anchors, resulting in the overestimation of fractional ratings. When Arabic numerals are used, the magnitude judgment of fractional ratings is affected by left-digit anchoring that results in the underestimation of the ratings. We believe that our contingent anchoring framework can be used to explain how the visual processing of symbols influence the magnitude perception of fractional numbers, predicting when fractional numbers will be overestimated or underestimated.

Theoretical Implications

Other than the main contributions discussed previously, our findings make several tangential yet noteworthy contributions to the literature at the intersection of visual perception and numerical cognition. First, while prior studies have indicated a human propensity for processing visual object representations more effectively than verbal ones (Lieberman and Culpepper 1965; Paivio, Rogers, and Smythe 1968), our findings present a nuanced view in the context of numerical quantities. Specifically, our Study 1 demonstrates that verbal

representations, such as Arabic numerals, yield more accurate estimations of fractional ratings compared to graphical representations like star symbols. This is further underscored by the observation that participants, when presented with both types of symbols, predominantly relied on Arabic numerals for magnitude estimation. These results challenge the conventional wisdom and suggest that in certain scenarios, particularly involving numerical information, verbal representations may lead to more efficient processing than graphical representations.

Furthermore, our findings provide new insights into the left-digit bias, particularly regarding its magnitude across different fractional values. While previous research has shown that fractional numbers (e.g., 3.99) are underestimated, it has not examined how the distance of the number from the left-digit influences this bias. That is, previous research suggests that the perceived magnitude of 2.25, 2.50, 2.75, and 2.99 will be anchored on the 2, the left-most digit. But will the bias in magnitude judgments vary across these four numbers? The extant research is silent on this issue. Our research demonstrates for the first time that the left-digit bias is more pronounced for values that are farther from the left-digit. Hence, the underestimation of magnitudes of fractional numbers will be the strongest for 2.99 and the weakest for 2.25. Stated differently, the extent of the left-digit bias in magnitude judgments is proportional to the distance between the left-most digit and the actual stimulus. The greater this distance, the larger is the bias. This finding significantly advances our knowledge of how the left-digit bias operates.

Practical Implications

The insights from our study provide crucial guidance for managers regarding the selection of rating symbols. Using graphical symbols instead of Arabic numerals may boost consumer perceptions of product quality, potentially leading to higher sales conversion rates,

decreased cart abandonment, and increased product expectations. Yet, our research reveals that when star ratings and Arabic numerals are displayed concurrently, consumers primarily focus on the Arabic numerals. This suggests that Arabic numerals may be more familiar or easier for consumers to interpret. Furthermore, our Studies 1 and 3 show that magnitude judgments for fractional ratings are consistently misestimated, regardless of the rating symbol used—graphical symbols tend to be overestimated, while Arabic numerals are underestimated on average. To ensure that retailers using graphical symbols do not have an unfair advantage over those using Arabic numerals, our findings advocate for the standardization of rating formats by either industry associations or regulatory bodies.

The implications of our research extend well beyond product ratings to encompass a wide range of applications where graphical symbols are used to convey numerical information. These include areas as varied as mobile battery indicators, vehicle fuel gauges, activity tracking on smartwatches, energy efficiency ratings, and vehicle safety ratings. Our finding that graphical representations can lead to the overestimation of numerical magnitudes is valuable for the design and interpretation of graphical displays across multiple sectors, including technology, transportation, and consumer electronics. We hope that our findings prompt future researchers to examine how various forms of graphical representations influence magnitude perceptions and downstream consumer behavior.

Limitations and Future Research

We have identified several areas that future research could explore to deepen our understanding of the influence of rating formats on consumer perception. First, while our research uses controlled experiments to study the impact of rating symbols on rating perception,

the complex dynamics of real online marketplaces, where multiple factors interplay in product comparisons and ratings, remain unaddressed. An important aspect not covered in our experiments is direct product comparison with different ratings. To robustly validate our findings, field studies in these multifaceted shopping environments are crucial, as they would provide insights into the natural interplay between various products and their ratings. Second, while our research shows overestimation of graphical ratings when the ratings were presented using discrete shapes (e.g., coloring three and a half stars out of five), we have not explored if this effect persists if the rating is presented using a continuous shape (like proportionately filling a single circle or a bar to represent the rating magnitude). Future research could investigate if the observed overestimation persists or varies when employing a continuous representation using a single shape and how it compares to the overestimation with discrete shapes.

Moreover, while our findings lean towards the anchoring and insufficient adjustment theory to explain how people interpret fractional ratings, we cannot say for sure that this is the only explanation. Alternative explanations, such as a cognitive weighing function where the round number value is disproportionately emphasized over the fractional value, are also consistent with our data. This opens the door for future research to dive deeper into these possibilities, helping to clarify how exactly consumers process and interpret numerical ratings.

Lastly, our study has focused on visual and verbal rating formats. However, we often share rating information about products through oral communication, like when we recommend a product to a friend. Thus, an interesting extension would be to explore the auditory representation of rating as a third modality. Investigating how numerical magnitudes are perceived and estimated through auditory senses could reveal unique patterns, contributing to our broader understanding of multimodal rating perception.

REFERENCES

- Bagchi, Rajesh and Amar Cheema (2013), “The Effect of Red Background Color on Willingness-to-Pay: The Moderating Role of Selling Mechanism,” *Journal of Consumer Research*, 39 (5), 947–60.
- Bagchi, Rajesh and Xingbo Li (2011), “Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity,” *Journal of Consumer Research*, 37 (5), 888–901.
- Barasz, Kate, Leslie K. John, Elizabeth A. Keenan, and Michael I. Norton (2017), “Pseudo-set Framing,” *Journal of Experimental Psychology: General*, 146 (10), 1460.
- Bhattacharya, Utpal, Craig W. Holden, and Stacey Jacobsen (2012), “Penny Wise, Dollar Foolish: Buy–Sell Imbalances on and Around Round Numbers,” *Management Science*, 58 (2), 413–31.
- Boegershausen, J., Yi, S., Cornil, Y., & Hardisty, D. J. (2023, working paper), “Testing the Digital Frontier: Opportunities and Validity Trade-offs in Digital Quasi-Experiments.”
- Booth, Julie L. and Robert S. Siegler (2008), “Numerical Magnitude Representations Influence Arithmetic Learning,” *Child Development*, 79 (4), 1016–31.
- Chen, Zoey and Nicholas H. Lurie (2013), “Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth,” *Journal of Marketing Research*, 50 (4), 463–76.
- Chevalier, Judith A. and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research*, 43 (3), 345–54.
- Dehaene, Stanislas (1997), *The Number Sense*, Oxford Univ. Press, Oxford.

- Dehaene, Stanislas, Emmanuel Dupoux, and Jacques Mehler (1990), “Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison.,” *Journal of Experimental Psychology: Human Perception and performance*, 16 (3), 626.
- Dehaene, Stanislas, Véronique Izard, Elizabeth Spelke, and Pierre Pica (2008), “Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures,” *Science*, 320 (5880), 1217–20.
- Epley, Nicholas and Thomas Gilovich (2004), “Are Adjustments Insufficient?,” *Personality and Social Psychology Bulletin*, 30 (4), 447–60.
- Epley, Nicholas and Thomas Gilovich (2006), “The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient,” *Psychological Science*, 17 (4), 311–18.
- Fisher, Matthew, George E. Newman, and Ravi Dhar (2018), “Seeing Stars: How the Binary Bias Distorts the Interpretation of Customer Ratings,” *Journal of Consumer Research*, 45 (3), 471–89.
- Foley, Mary Ann, Hugh J. Foley, Francis T. Durso, and N. Kyle Smith (1997), “Investigations of Closure Processes: What Source-Monitoring Judgments Suggest About What is ‘Closing,’” *Memory & cognition*, 25 (2), 140–55.
- Foley, Mary Ann, Hugh J. Foley, Rachel Scheye, and Angelica M. Bonacci (2007), “Remembering More Than Meets the Eye: A Study of Memory Confusions about Incomplete Visual Information,” *Memory*, 15 (6), 616–33.
- Gerbino, Walter (2020), “Amodal Completion Revisited,” *i-Perception*, 11 (4), 2041669520937323.

- Gigerenzer, Gerd and Ulrich Hoffrage (1999), “Overcoming Difficulties in Bayesian Reasoning: A Reply to Lewis and Keren (1999) and Mellers and McGraw (1999).”
- Ginzberg, Eli (1936), “Customary Prices,” *The American Economic Review*, 26 (2), 296–296.
- Griffin, Dale and Amos Tversky (1992), “The Weighing of Evidence and the Determinants of Confidence,” *Cognitive Psychology*, 24 (3), 411–35.
- Hagtvedt, Henrik and S. Adam Brasel (2017), “Color Saturation Increases Perceived Product Size,” *Journal of Consumer Research*, 44 (2), 396–413.
- Hayes, Andrew F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, Guilford publications.
- Hinrichs, James V., Dale S. Yurko, and Jing-Mei Hu (1981), “Two-Digit Number Comparison: Use of Place Information.,” *Journal of Experimental Psychology: Human Perception and Performance*, 7 (4), 890.
- Jia, He (Michael), Echo Wen Wan, and Wanyi Zheng (2023), “Stars versus Bars: How the Aesthetics of Product Ratings ‘Shape’ Product Preference,” *Journal of Consumer Research*, ucac043.
- Jiang, Zhenling (2022), “An Empirical Bargaining Model with Left-Digit Bias: A Study on Auto Loan Monthly Payments,” *Management Science*, 68 (1), 442–65.
- Kanizsa, Gaetano (1979), *Organization in Vision: Essays on Gestalt Perception*, Praeger Publishers.
- King, Dan and Chris Janiszewski (2011), “The Sources and Consequences of the Fluent Processing of Numbers,” *Journal of Marketing Research*, 48 (2), 327–41.

- Krishna, Aradhna (2006), “Interaction of Senses: The Effect of Vision Versus Touch on the Elongation Bias,” *Journal of Consumer Research*, 32 (4), 557–66.
- Kyung, Ellie J, Manoj Thomas, and Aradhna Krishna (2017), “When Bigger Is Better (and When It Is Not): Implicit Bias in Numeric Judgments,” *Journal of Consumer Research*, 44 (1), 62–79.
- Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor (2012), “Heuristic Thinking and Limited Attention in the Car Market,” *American Economic Review*, 102 (5), 2206–36.
- de Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016a), “Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings,” *Journal of Consumer Research*, 42 (6), 817–33.
- de Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016b), “Star Wars: Response to Simonson, Winer/Fader, and Kozinets,” *Journal of Consumer Research*, 42 (6), 850–57.
- Lembregts, Christophe and Jorge Pena-Marin (2021), “Numbers and Units Affect Goal Pursuit Organization and Motivation,” *Journal of Consumer Psychology*, 31 (1), 37–54.
- Lieberman, Lewis R. and James T. Culpepper (1965), “Words versus Objects: Comparison of Free Verbal Recall,” *Psychological Reports*, 17 (3), 983–88.
- van Lier, R. J. and W. Gerbino (2015), “Perceptual Completions in Oxford Handbook of Perceptual Organization (ed. Wagemans, J.) 294–320,” Oxford University Press.
- Manning, Kenneth C. and David E. Sprott (2009), “Price Endings, Left-Digit Effects, and Choice,” *Journal of Consumer Research*, 36 (2), 328–35.

- McKinsey (2021), “Five-Star Growth: Using Online Ratings to Design Better Products | McKinsey,” (accessed February 11, 2024), [available at <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/five-star-growth-using-online-ratings-to-design-better-products?cid=soc-web>].
- Monga, Ashwani and Rajesh Bagchi (2012), “Years, Months, and Days versus 1, 12, and 365: The Influence of Units versus Numbers,” *Journal of Consumer Research*, 39 (1), 185–98.
- Olsen, Asmus Leth (2013), “Leftmost-Digit-Bias in an Enumerated Public Sector? An Experiment on Citizens’ Judgment of Performance Information.,” *Judgment & Decision Making*, 8 (3).
- Paivio, Allan, Timothy B. Rogers, and Padric C. Smythe (1968), “Why are Pictures Easier to Recall than Words?,” *Psychonomic Science*, 11 (4), 137–38.
- Pandelaere, Mario, Barbara Briers, and Christophe Lembregts (2011), “How to Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons,” *Journal of Consumer Research*, 38 (2), 308–22.
- Pena-Marin, Jorge and Rajesh Bhargave (2016), “Lasting Performance: Round Numbers Activate Associations of Stability and Increase Perceived Length of Product Benefits,” *Journal of Consumer Psychology*, 26 (3), 410–16.
- Pessoa, Luiz and Peter De Weerd (2003), *Filling-In: From Perceptual Completion to Cortical Reorganization*, Oxford University Press.
- Pessoa, Luiz, Evan Thompson, and Alva Noë (1998), “Finding Out About Filling-In: A Guide to Perceptual Completion for Visual Science and the Philosophy of Perception,” *Behavioral and Brain Sciences*, 21 (6), 723–48.

- PowerReviews (2021), “Survey: The Ever-Growing Power of Reviews,” *PowerReviews*, (accessed February 11, 2024), [available at <https://www.powerreviews.com/insights/power-of-reviews-survey-2021/>].
- Raghubir, Priya and Aradhna Krishna (1999), “Vital Dimensions in Volume Perception: Can the Eye Fool the Stomach?,” *Journal of Marketing research*, 36 (3), 313–26.
- Ramachandran, Vilayanur S. (1992), “Filling in Gaps in Perception: Part I,” *Current Directions in Psychological Science*, 1 (6), 199–205.
- Rensink, Ronald A. and James T. Enns (1998), “Early Completion of Occluded Objects,” *Vision Research*, 38 (15–16), 2489–2505.
- Rosch, Eleanor (1975), “Cognitive Reference Points,” *Cognitive Psychology*, 7 (4), 532–47.
- Rozenkrants, Bella, S Christian Wheeler, and Baba Shiv (2017), “Self-Expression Cues in Product Rating Distributions: When People Prefer Polarizing Products,” *Journal of Consumer Research*, 44 (4), 759–77.
- Santana, Shelle, Manoj Thomas, and Vicki G. Morwitz (2020), “The Role of Numbers in the Customer Journey,” *Journal of Retailing*, 96 (1), 138–54.
- Schley, Dan R. and Ellen Peters (2014), “Assessing ‘economic value’: Symbolic-number mappings predict risky and riskless valuations,” *Psychological Science*, 25 (3), 753–61.
- Schoenmueller, Verena, Oded Netzer, and Florian Stahl (2020), “The Polarity of Online Reviews: Prevalence, Drivers and Implications,” *Journal of Marketing Research*, 57 (5), 853–77.

- Siegler, Robert S. and John E. Opfer (2003), “The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity,” *Psychological Science*, 14 (3), 237–50.
- Stiving, Mark and Russell S. Winer (1997), “An Empirical Analysis of Price Endings with Scanner Data,” *Journal of Consumer Research*, 24 (1), 57–67.
- Thomas, Manoj, Daniel H. Simon, and Vrinda Kadiyali (2010), “The Price Precision Effect: Evidence from Laboratory and Market Data,” *Marketing Science*, 29 (1), 175–90.
- Thomas, Manoj and Vicki Morwitz (2005), “Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition,” *Journal of Consumer Research*, 32 (1), 54–64.
- Thomas, Manoj and Vicki Morwitz (2009), “Heuristics in Numerical Cognition: Implications for Pricing,” in *Handbook of pricing research in marketing*, Edward Elgar Publishing, 132–49.
- To, Rita Ngoc and Vanessa M. Patrick (2021), “How the eyes connect to the heart: The influence of eye gaze direction on advertising effectiveness,” *Journal of Consumer Research*, 48 (1), 123–46.
- Townsend, Claudia and Barbara E. Kahn (2014), “The ‘Visual Preference Heuristic’: The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload,” *Journal of Consumer Research*, 40 (5), 993–1015.
- Tversky, Amos and Daniel Kahneman (1974), “Judgment Under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking Under Uncertainty.,” *Science*, 185 (4157), 1124–31.

Watson, Jared, Anastasiya Pocheptsova Ghosh, and Michael Trusov (2018), “Swayed by the Numbers: The Consequences of Displaying Product Review Attributes,” *Journal of Marketing*, 82 (6), 109–31.

Yan, Dengfeng and Jaideep Sengupta (2021), “The Effects of Numerical Divisibility on Loneliness Perceptions and Consumer Preferences,” *Journal of Consumer Research*, 47 (5), 755–71.

CHAPTER 2

YEAR VERSUS LENGTH:

THE INFLUENCE OF TIME FRAMING ON PRODUCT VALUATION

Abstract

This research demonstrates that time framing influences product evaluation when the temporal attribute guides the product's perceived quality and value. For instance, analysis of an archival whiskey auction dataset revealed that whiskeys marketed with length framing (such as "12-year-old whiskey") fetched higher prices than those with year framing (such as "2012-whiskey"). Conversely, for second-hand goods, such as used furniture, year framing (e.g., "bought in 2022") increases consumer willingness to pay compared to length framing (e.g., "bought 2 years ago"). This effect extends to products with future temporal attributes, affecting the valuation of financial instruments such as certificates of deposit. The observations are due to changes in subjective time perception: time framed by its length (e.g., "10 years later/ago") is perceived as more distant compared to year framing (e.g., "2034 / 2014"). Owing to the logarithmic scale of number perception, where larger numbers are perceived as being closer together, numerical changes feel smaller in year framing (from 2024 to 2034) compared to length framing (from 0 to 10), consequently leading to shorter perceived time intervals. This effect of time framing has critical implications for how consumers evaluate products with a time component, weigh time against money, and make investment decisions.

Which whiskey would you value more: a ‘12-year-old whiskey’ or a ‘2012 whiskey’? Are people more likely to save money for retirement when the timing of the need is framed as ‘35 years later’ rather than ‘in 2059’? Should a seller list her used furniture as ‘a 5-year-old mahogany bed’ or ‘a mahogany bed purchased in 2019’? These questions highlight a common thread: the framing of time intervals.

Although the choice of time description format may sometimes be determined by context, norms, or practicality, it is often arbitrary. For example, both formats are feasible and commonly used when listing used goods. We find that the way we communicate time intervals, either by the number of intervening years (e.g., 5 years ago or 5 years later) or by the boundary year (e.g., 2019 or 2029), can substantially influence our perception of time. This, in turn, affects evaluations when time is an important factor in determining product value. This research aims to examine whether describing time intervals in terms of their boundary year, rather than their length, alters people’s product evaluations and perceptions of the length of time. Additionally, we investigate the underlying reasons for this effect.

We conducted a series of experiments to investigate how the framing of time intervals influences consumers’ evaluative judgments. In our experiments, we presented the same time interval in two ways: either by using the boundary year (e.g., 2020 or 2027) or by using the length of the interval (e.g., 4 years ago or 3 years later). Our results indicate that year framing leads people to perceive the time interval as shorter compared to length framing, regardless of whether the time interval refers to the past or the future. The choice of time framing can significantly impact market decisions, such as evaluations of second-hand products, whiskeys, and investment returns.

Importantly, we examined the cognitive processes underlying this effect and found that

subjective time perception is influenced by the nonlinear nature of mental magnitude representations. We propose that in contexts where time is described by years (e.g., from 2024 to 2012), the perceived change feels smaller due to the logarithmic nature of the mental number line, where larger numbers are perceived as closer together. Conversely, when time is described by length (e.g., 12 years ago), the change feels more substantial because each unit represents a more significant portion of the mental number line.

Our research has significant implications for three key areas of consumer behavior: magnitude judgment, time perception, and numerical cognition. It contributes to the growing body of literature exploring how specifying an attribute in an alternative unit can influence magnitude judgments, as demonstrated in prior research (Burson, Larrick, and Lynch 2009; LeBoeuf 2006; Monga and Bagchi 2012; Munichor and LeBoeuf 2018; Pandelaere, Briers, and Lembregts 2011; Sokolova 2023; Zhang and Schwarz 2012). Specifically, our study focuses on time magnitude judgments. For example, Sokolova (2023) showed that using days-of-the-week information in descriptions of temporal intervals (e.g., “from Monday, February 1 to Saturday, February 6” instead of “from February 1 to February 6”) can make people perceive the time interval as longer due to shifts in their implicit scales.

Our research offers significant practical applications, particularly in the context of duration estimation, which is integral to numerous daily decisions. These decisions encompass future planning, goal setting, time and money trade-offs, and product evaluations. Understanding which time framing leads to the relative underestimation of duration, and why, can guide policymakers and marketing professionals in ways that can benefit both consumer and company welfare. For instance, the popular whiskey brand Macallan employs length-based framing to market some whiskeys, such as “The Macallan Double Cask 18 Years Old,” and year-based

framing for others, like “The Macallan Fine & Rare 1991 Vintage.” Our findings indicate that this choice of framing can significantly impact consumer perception and purchase decisions.

The remainder of this article is organized as follows. We begin by reviewing relevant research on time perception and numerical cognition, which forms the foundation for our hypotheses regarding the impact of year-length framing on perceived lengths of time intervals and evaluative judgments. Next, we will present the results from a series of studies designed to test these hypotheses.

THEORETICAL BACKGROUND

Time Interval Framing

Research indicates that the way time intervals are described can significantly influence people's perceptions, judgments, and decisions. One way to vary interval descriptions is by changing the units used. For instance, 1 year can be expressed as 12 months, 52 weeks, or 365 days. Studies have shown that when larger units, such as weeks, are used to describe time intervals, they are perceived as longer than when smaller units, such as days, are used. This phenomenon is known as the “unitosity effect.” Monga and Bagchi (2012) found that people perceived a change in delivery time from 1 to 3 weeks as longer than a change from 7 to 21 days, because larger units are typically associated with longer time periods.

Similarly, Ülkümen and Thomas (2013) found that consumers were less likely to adopt a one-year diet plan compared to a 12-month plan, as the one-year plan seemed longer and more challenging. In addition to the unitosity effect, biases in time distance perception can also be influenced by the “numerosity effect.” The numerosity effect suggests that as the number increases, perceptions of an interval's length may also increase. For example, Zhang and Schwarz (2012) discovered that when time intervals were presented with smaller units and larger numbers

(e.g., 120 minutes instead of 2 hours), they were perceived as more precise. Similarly, Pandelaere, Briers, and Lembregts (2011) found that people were more sensitive to differences between products when smaller units with larger numbers were used (e.g., warranty expressed in months) compared to larger units with smaller numbers (e.g., warranty expressed in years).

Another way to vary time interval descriptions is by using dates (e.g., October 17) or equivalent duration of time (e.g., 120 days, 16 weeks, 4 months) to describe the interval. Studies have shown that people tend to discount future monetary rewards less when the time is described using dates than when it is described using a duration of time (LeBoeuf 2006; Malkoc, Zauberman, and Bettman 2010; Read et al. 2005). This might be because duration descriptions make the time interval seem subjectively longer (LeBoeuf 2006). Research has also demonstrated that the format used (date vs. duration) can influence the shape of the discounting function. Specifically, when a duration format is employed, the discounting function tends to be hyperbolic, whereas this hyperbolic pattern is absent when the date format is used (Malkoc, Zauberman, and Bettman 2010; Read et al. 2005). Furthermore, consistent with endpoint framing leading to lower future discount rates than length framing, Frederick, Read, and LeBoeuf (2008) found that participants' discount rates were lower when the future time was described using the corresponding age (e.g., "when you are 47") than when described using the number of years (e.g., "10 years").

In summary, previous research has consistently shown that individuals discount future transactions less when time intervals are specified using dates rather than durations. However, the proposed underlying processes vary and remain untested. One proposed explanation is a preference for precision, where dates provide a specific point in time, making people more certain they will receive the reward, thus increasing patience (Read et al. 2005). This, along with some other explanations (see Table 2.1 for a summary of mechanisms explored by previous research), is

unique to intertemporal choice with discounting of future outcomes and cannot explain whether and how year-length framing influences product evaluations such as buying products with a temporal attribute that affects value, like whiskey or used furniture.

Table 2.1: Mechanisms Explored in Prior Research on Date-Duration Framing Effects

S. No.	MECHANISM	SUMMARY
1	Preference for precision (Read et al. 2005; Zhang and Schwarz 2012)	Read and colleagues (2005) find that discount rates for future outcomes depending on whether time is described using dates or in terms of duration. One explanation they proposed for this phenomenon is the preference for precision hypothesis, which suggests that individuals tend to prefer the precision of a specific date over the vagueness of a duration, leading to more patient behavior when waiting for a specific date due to the perceived certainty of receiving the reward on the specified date. This hypothesis has also been supported by Zhang and Schwarz (2012) in the context of consumer trust in marketing communications.
2	Abstract and concrete representation (Malkoc, Zauberman, and Bettman 2010)	When it comes to future time discounting, the use of duration (vs. date) frames can increase people's tendency to discount future rewards. Duration frames, which describe the task as a departure from the present and emphasize the length of time one needs to wait for a future outcome, tend to elicit concrete representations. In contrast, date frames shift the focus to the future moment of the outcome without emphasizing the length of time one needs to wait, leading to more abstract representations. The abstract (vs. concrete) mindset induced by the date (vs. duration) framing results in lower discounting of future outcomes.
3	Vividness or accessibility (LeBoeuf 2006; Read et al. 2005)	According to LeBoeuf (2006), when considering future time discounting, duration frames explicitly highlight the time required to obtain a future reward, while date framing omits this detail and may not even prompt calculation. As a result, consumers are more likely to focus on the interval length when presented with duration framing, making the passage of time more vivid and noticeable, and ultimately causing the interval to seem longer. Conversely, date descriptions may direct attention towards other factors, such as the final transaction, resulting in less emphasis on the interval length. Read and colleagues (2005) also suggested that, compared to duration framing, date framing could lead to underestimating the temporal interval.
4	Relative attention on value (Read et al. 2005)	The framing of time as duration (vs. date) in discounting tasks can affect how much attention is given to the value compared to the timing of a future outcome. In this context, date framing tends to emphasize the value more than duration framing does. Consequently, larger rewards may be preferred even if they are received later, as the focus remains on the outcome value.
5	Different decision strategy (Read et al. 2005)	In discounting tasks, framing time as duration (versus date) can either aid or hinder various decision-making strategies. When presented with duration, respondents may be more inclined to use a compensatory strategy, as both duration and amounts are continuous numeric variables, which can facilitate evaluation and choice between outcomes. However, date descriptions can obstruct this process. In this case, participants must either convert dates into corresponding delays or employ a non-compensatory decision strategy, such as consistently selecting the option that excels in the most important attribute.
6	Similarity of time points (Read et al. 2005)	Previous studies have suggested that discounting decreases in relation to the perceived similarity between the time-points marking the beginning and end of an interval (Leland 2002). Following this idea, the similarity hypothesis proposes that the duration/date effect emerges in discounting tasks because two dates might appear more similar than two durations covering the same time period.

7	Anchoring and adjustment (LeBoeuf and Shafir 2009)	In this study, the authors had participants estimate the duration of uncertain time intervals using either duration-based (e.g., “How many days until...”) or date-based (e.g., “On what date...”) formats. The results demonstrated that when estimating the number of days, participants consistently underestimated the duration of these intervals compared to estimating a specific date. The authors attributed this to an anchoring and adjustment process, where individuals base their estimates on the present moment and adjust incrementally according to the unit, leading to insufficient adjustment and systematic underestimation. Conversely, date-based estimates involve more global processing without any anchor.
8	Relative focus on outcome vs. process of goal (Munichor and LeBoeuf 2018)	The authors discovered that people are more likely to pursue their goals when the completion interval is described in terms of duration (e.g., “two weeks from now”) rather than a specific date (e.g., “between today and November 17”). They propose that dates make it easier for individuals to recall other obligations during that time, causing them to concentrate more on the process of pursuing the goal, which can be unpleasant. On the other hand, durations present the interval in isolation, enabling individuals to focus more on the positive outcome of accomplishing their goal.

NOTE.— The mechanisms listed above were either indirectly observed or untested in the referenced studies, with two exceptions. First, Munichor and LeBouef (2018) demonstrated the ‘outcome vs. process focus’ mechanism, although its applicability to non-goal domains is limited. Second, LeBoeuf (2006) established the main effect of duration (vs. date) framing on subjective interval length. Importantly, this table excludes mechanisms that have been previously ruled out by research (e.g., the subadditivity mechanism in LeBoeuf’s (2006) Study 5). Although this list is not comprehensive, no known studies have explored the logarithmic scale account as an explanation for time framing effects.

Broadening our exploration beyond intertemporal choice and discounting domains while maintaining a focus on date versus duration framing, past research has found that people tend to underestimate the time required to complete a future task when using a duration rather than a date. LeBoeuf and Shafir (2009) proposed that this is due to the “anchoring and insufficient adjustment” bias. They suggested that when people make estimates using the number of days, they tend to anchor on the present moment and make insufficient adjustments from that anchor, resulting in lower estimates. Another area where date-duration framing affects consumer behavior is goal setting. Munichor and LeBoeuf (2018) found that consumers are more likely to adopt goals, such as going on a diet, when the time interval is described as a duration, such as “two weeks from now,” rather than a date, such as “July 3.” This is because date descriptions can elicit competing obligations within the specified time frame, whereas a duration format presents the time interval

in isolation. Nevertheless, the processes of anchoring, adjustment, and competing obligations cannot predict whether and how year-length framing influences product evaluations.

Our theory on numerical change perception highlights the psychological differences between year-length effect and date-duration framing effects that were observed in previous research. In year-length framing, year framing involves much larger nominal numbers (e.g., 2029) compared to length framing (e.g., 5 years), which is crucial to our theory. Conversely, in date-duration framing, the opposite might be true: dates may involve smaller numbers than durations. For instance, when comparing ‘October 17’ to ‘120 days’, 17 is much smaller than 120. Dates also tend to combine numbers and words⁹, while years are purely numerical, making our theory on the nonlinear mapping of large numbers difficult to apply to differences between dates and durations. These distinctions suggest that date-duration and year-length effects are psychologically distinct. The next section further elaborates on our theory of numerical change perception.

Analog Model of Numerical Cognition

While prior work on time perceptions using different interval formats has proposed a variety of possible mechanisms (see Table 2.1), our focus on using intervals with equivalent precision and only numerical values with units (e.g., 2020 versus 4 years ago) allows us to narrow our proposed mechanism to one of numerical perception, and specifically the analog model of numerical perception. This model provides a foundational understanding of how numeric symbols are internally processed. According to this model, individuals translate numeric symbols into estimated magnitudes along a conceptual mental number line—a process known as mapping symbolic representations to analog magnitudes (Dehaene 1997; Dehaene, Dupoux, and Mehler

⁹ Dates are expressed either using a combination of numbers and words (e.g., 17 October 2024) or using numbers and symbols (e.g., 10/17/2024; 10-17-2024). Regardless of the format, dates do not function as numbers.

1990). This mental number line is primarily located within the inferior parietal cortex, particularly the horizontal intraparietal sulcus, which is activated by various numeral formats, including Arabic, spelled-out, and spoken numerals (Dehaene et al. 2003; Pinel et al. 2001). The involvement of this brain area suggests it plays a pivotal role in the abstract representation and manipulation of numeric information, enabling cross-modal interpretations of numbers. Moreover, humans and other animals demonstrate intuitive, nonverbal representations of numerosity that activate automatically in response to numeric stimuli, allowing them to perform a range of numerical tasks from simple arithmetic to more complex mathematical operations (Barth et al. 2006; Brannon and Terrace 1998). Extensive research indicates that the processing of numeric magnitudes by humans and various laboratory animals mirrors the processing of other physical magnitudes, adhering to established psychophysical laws.

Consistent with psychophysical laws, individuals exhibit a numeric distance effect, where far-apart values are discriminated more easily than closer values, and a size effect, where the ability to distinguish magnitudes decreases as the magnitudes increase (Moyer and Landauer 1967; Parkman 1971). These discriminability effects are believed to originate from the inexact mappings of symbolic numbers to mental magnitudes (Dehaene et al. 2008; Feigenson, Dehaene, and Spelke 2004; Halberda, Mazocco, and Feigenson 2008).

Further investigations into the mental representation of numbers reveal that the internal mental number line is logarithmic rather than linear (Dehaene 2001; Dehaene et al. 2008). This logarithmic nature means that larger numbers appear closer together than smaller numbers. This concept is supported by Weber and Fechner's psychological theories, which suggest that the ability to differentiate between two stimuli—such as sound, light, or temperature—is influenced by the magnitude of the initial stimuli (Fechner 1860; Weber 1834). Correspondingly, the mental number

line exhibits diminishing discriminability with increasing numbers. For instance, the difference between 10 and 20 feels more significant than the difference between 110 and 120, even though both intervals are numerically equivalent. This nonlinear perception impacts not only fundamental numerical cognition but also complex decision-making processes in daily life.

In consumer behavior, the nonlinear configuration of the mental number line significantly influences how numerical information, such as prices, is perceived. For instance, a \$5 discount on a \$50 purchase (i.e., the difference between \$50 and \$45) is perceived as more significant than the same \$5 discount on a \$1000 purchase (i.e., the difference between \$1000 and \$995). This reflects the impact of nonlinear number perception and proportionate thinking in deal evaluations (Thomas and Morwitz 2009a; b). The nonlinear mapping of symbolic numbers to mental magnitudes also underpins fundamental economic behaviors such as diminishing marginal utility (DMU). In DMU, the perceived value from incremental increases diminishes as the quantity increases. Researchers suggest that this behavioral pattern is at least partially driven by the nonlinear mapping of symbolic numbers onto mental magnitudes (Schley and Peters 2014).

Proposed Year-Length Effect

The theoretical framework for our investigation into the impact of year-length framing on subjective time perception is founded on two core postulates. First, the perception of temporal distance—whether framed as ‘the year 2012’ or ‘12 years ago’—is influenced by the nominal values used in different framings: ‘2012’ versus ‘12’. These values prompt distinct numerical comparisons, with year framing (e.g., ‘2012 and 2024’) involving larger numbers, and length framing (e.g., ‘12 units from now’) involving much smaller numbers. These numerical comparisons guide subjective time perception. Second, perceptions of numerical changes are

shaped by the nonlinear, or logarithmic, nature of the mental number line. This logarithmic shape causes larger numbers to be perceived as closer together than smaller numbers, thereby affecting how temporal distances are evaluated based on the framing used. As a result, temporal intervals described in year framing are perceived as shorter (i.e., closer together) compared to those described in length framing, even though both represent the same amount of time. We posit that these differences will influence how people evaluate products and perceive time intervals, ultimately affecting their decisions and judgments (see Figure 1.1 for a graphical representation of our conceptual framework). Based on this theoretical framework, we propose the following hypotheses:

H1: Products with a time attribute where older is better (e.g., whiskey) will be evaluated less favorably when the time is described using year framing (e.g., 2012 whiskey) compared to length framing (e.g., 12-year-old whiskey).

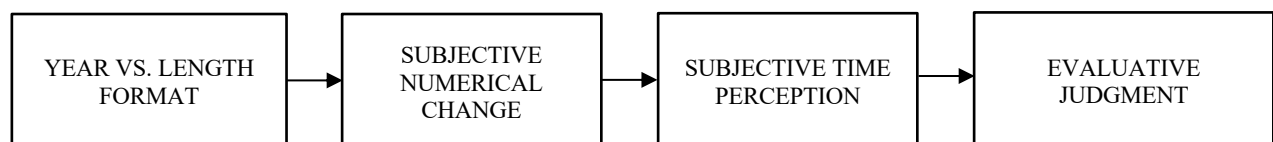
H2: Products with a time attribute where newer is better (e.g., secondhand furniture) will be evaluated more favorably when the time is described using year framing (e.g., a sofa bought in 2022) compared to length framing (e.g., a sofa bought 2 years ago).

H3: Time intervals described using year framing will be perceived as subjectively shorter than equivalent time intervals described using length framing, applicable to both past and future intervals.

H4a: The effect of year-length framing on time perception is driven by the subjective numerical change perception in the nominal values of the time intervals.

H4b: The effect of year-length framing on time perception will attenuate when the nominal values of time intervals are represented on a linear scale.

Figure 2.1: Drivers of Year-Length Effect



Empirical Package

We examined the year-length effect on consumer evaluative judgments through an empirical package that included an archival dataset from whiskey auctions, comprising 50,916 entries, and eight laboratory studies with a total of 6,224 participants. All experimental studies were preregistered on aspredicted.org, and the data along with surveys are available in the Research Box (<http://bit.ly/3GXehUX>). Supplementary materials can also be found in the Chapter 2 Appendix. Studies 1, 2, 3A, and 3B establish the year-length effect. The remaining studies focus on process: Studies 4 and 5 explore how time framing influences subjective time perception, while Studies 6A and 6B demonstrate the role of numerical change perception in driving this effect.

STUDY 1: ACTUAL WHISKEY AUCTIONS

This study aims to examine the economic impact of whiskey age representation—whether presented in terms of year or length—on consumer purchasing decisions at auctions. To do this, we analyzed actual whiskey auction transactions. Our theory predicts that whiskeys that use the year (vs. length) framing to specify their age would be perceived as less old, and hence less valuable.

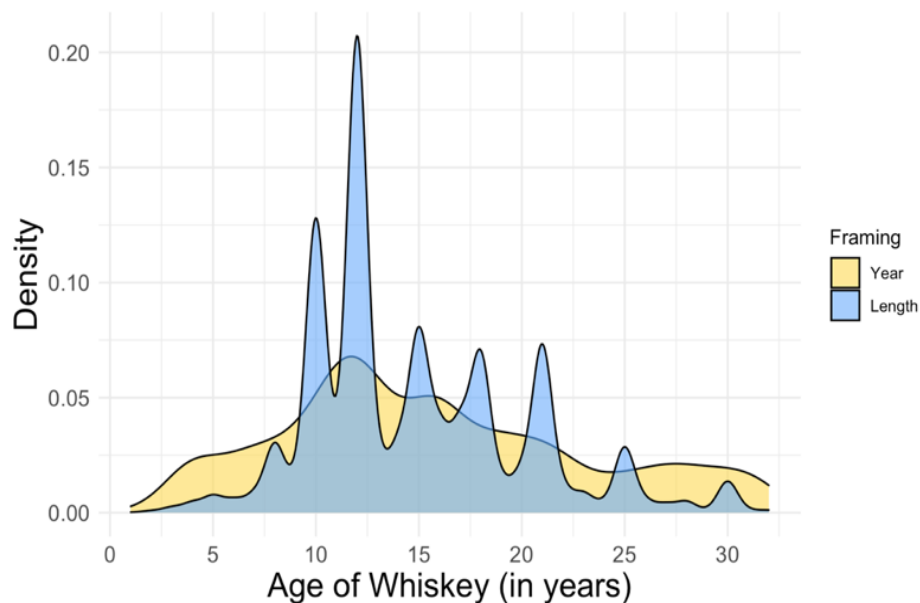
We selected whiskeys to examine the impact of time framing because, unlike wine, an unopened bottle of whiskey does not improve with age (Teeter 2020). When stored correctly, the whiskey in a bottle remains inert and experiences minimal change over time. As a result, marketers need not update the printed age on the bottle, which allows them to employ either time format—year or length—to name and label the whiskey.

This characteristic of whiskey ensures that a 10-year-old whiskey bottle will consistently remain ten years old, and similarly, a whiskey distilled (or manufactured) in 2014 and bottled in 2024 will perpetually be ten years old. Additionally, while the distiller can choose to highlight

either the year or length on the label, the distillation year, bottling year, and age are often still printed elsewhere on the bottle, along with various other details about the whiskey.

The age of whiskey is calculated by subtracting the year of manufacture from the bottled year. Whiskeys are typically sold either in the year they are bottled or within the following few years. For those labeled with ‘year framing’, the title includes the year of manufacture (e.g., 2005) to signify its age. Consequently, relying solely on the year of manufacture, without accounting for the bottled year (e.g., 2023), can lead to an overestimation of age for year-framed whiskeys as time from bottling increases, potentially resulting in higher bids. Therefore, our setting serves as a conservative test to our hypothesis.

Figure 2.2: Age Distribution of Whiskey by Framing



NOTE.— The figure displays the age distribution of auctioned whiskeys, with 42,992 auctions using length framing and 7,924 using year framing.

Data

Secondary Data. We compiled our dataset from secondary data by scraping the website of a leading UK-based whiskey auction company. This dataset encompasses extensive details on whiskey attributes such as age, brand, and final sale price. For the purposes of our study, we use

the term “whiskey” to include both the typically Irish or American “whiskey” and “whisky” from Scottish, Canadian, and Japanese origins.

This auction platform hosts monthly online auctions, attracting a global audience of buyers and sellers. Sellers must register and send their whiskey bottles to the company for authentication prior to listing. Each auction runs for a month and features hundreds of whiskeys available for bidding. Buyers worldwide can participate and place bids on their chosen bottles, facilitated by the auction company’s international shipping services. At the conclusion of each auction, the highest bidder is awarded the bottle. While our dataset provides extensive details on whiskey characteristics and auction outcomes, it lacks specific demographic information about the buyers and sellers. Nevertheless, the depth of the dataset in terms of whiskey details and transactional data offers a valuable opportunity to test our hypothesis in a meaningful real-world context.

Data Exclusions. To ensure the reliability of our analysis, we cleaned the dataset and applied the 1.5 Interquartile Range (IQR) rule to identify and exclude outliers in the (log) sold price and age of whiskeys. Entries exceeding 1.5 times the IQR below the first quartile or above the third quartile for these variables were removed to ensure robustness against extreme values. Refer to Figure 2.2 for distribution of age of whiskeys.

Moreover, we identified a subset of whiskeys ($n = 21,995$) labeled with both year and length terms in their age descriptions, such as ‘Macallan - 30 Year Old (1989)’. Initial observations suggested these could represent a third category of time framing. However, further analysis, detailed in the Chapter 2 Appendix, showed that bottle images often emphasized one type of framing over the other, complicating their straightforward classification. As a result, we excluded these whiskeys from the primary analysis. Interestingly, the effect of year length

framing persists even if these whiskeys are considered under a combined ‘both year and length’ category. We explore the implications of this dual framing further in Study 4 of this paper.

Our main analysis was conducted on the final dataset of 50,916 whiskey auctions from March 2016 to January 2024. The mean sold price of the whiskey in our data was £195 (~\$245), and the mean age of the whiskey was 15 years. The dataset contained independent variations across time framing (year vs. length), brand (1385 brands), type of whiskey (19 types such as single malt and bourbon whiskey), and region (eight different regions), thus making this an excellent dataset to study the relative effects of year and length framing on whiskey valuation. For detailed summary statistics across the various types and brands, refer to the Chapter 2 Appendix.

Key Variables. Our main dependent variable was the log-transformed sold price, representing the final auction price awarded to the highest bidder. It’s important to note that while our dataset did not include details on competing bids, the iterative nature of the bidding process (allowing bidders to increase their offers) and the month-long auction duration suggest that the final sold price was shaped by multiple bids, reflecting the collective valuation of whiskey enthusiasts.

Table 2.2: Descriptive Statistics for Whiskey Auction Data

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Size (in cl)</i>	<i>Strength (% ABV)</i>	<i>Mean Sold Time</i>	<i>Unique brands</i>	<i>N</i>
Year	£265	15.7	70	50%	mid 2020	466	7924
Length	£182	14.8	70	46%	mid 2020	1123	42992
Total	£195	14.9	70	47%	mid 2020	1385	50916

Our primary independent variable was the time framing used to describe the whiskey’s age. For example, ‘1989 The Macallan Fine & Rare Vintage Single Malt Scotch Whisky’ illustrates year framing, whereas ‘Macallan Highland Single Malt Scotch Whisky 30 Years Old’

exemplifies length framing. The dataset predominantly features length framing (84%), with year framing used in 16% of cases. Detailed images of whiskey bottles with varying framings are provided in the Chapter 2 Appendix. Descriptive statistics by time framing are presented in Table 2.2.

Other Predictors. In addition to sold price and time framing, our dataset includes a comprehensive range of whiskey characteristics. These encompass the brand name, whiskey age¹⁰, year of manufacture, type, region, distillery status at auction time, and a detailed text description. This description provides insights into the whiskey lot and distillery, highlighting aspects such as whiskey rarity, awards, and any pertinent warnings. Other physical attributes recorded include bottle size (in centiliters), delivery weight (in kilograms), individual bottle number, total number of bottles produced, lot number, and the alcohol strength (% ABV). Transactional details like VAT applicability and the date of sale are also captured. We added all these variables in our model.

Some whiskeys ($n = 4607$) on the auction platform were listed with a badge of VAT applicability. The presence of this badge meant that the bidding price would be subject to standard VAT at the time of checkout¹¹. Thus, we expect that consumers, to compensate, would systematically bid lower for a whiskey with the VAT badge compared to the identical whiskey without the badge. Therefore, we believe that the effect of the VAT applicability on the sold price can be used as a benchmark to evaluate the size of the time framing effect.

¹⁰ For length-framed whiskeys, we deduced age directly from length information. For year framed ones, we calculated age by computing the difference between bottled year and year of manufacture. The bottled year and year of manufacture were obtained from the whiskey description and the title, respectively.

¹¹ Without the VAT badge, whiskey is sold through auctioneers' VAT margin scheme, where VAT applies only to the auctioneer's fees, not the whiskey value. More on the UK's auctioneers' VAT margin scheme: <https://www.gov.uk/guidance/using-the-auctioneers-vat-margin-scheme>.

Table 2.3: OLS Regression Models for Whiskey Auctions

Predictor Variables	Dependent Variable: (log) Price		
	(1)	(2)	(3)
Framing [€] : Year	-.086***	-.085***	-.081***
	(-.10, -.07)	(-.10, -.07)	(-.10, -.06)
Age	.531***	.499***	.352***
	(.53, .54)	(.49, .51)	(.17, .53)
Size	.155***	.142***	.142***
	(.15, .16)	(.14, .15)	(.14, .15)
Strength	.387***	.354***	.344***
	(.38, .39)	(.35, .36)	(.34, .35)
Delivery weight	.156***	.152***	.152***
	(.15, .16)	(.15, .16)	(.15, .16)
VAT applicable [€] : Yes	-.085***	-.033**	-.043***
	(-.10, -.07)	(-.06, -.01)	(-.07, -.02)
Rarity descriptor: Very Rare [€]		.563***	.580***
		(.50, .62)	(.52, .64)
Rarity descriptor: Rare [€]		.337***	.320***
		(.31, .36)	(.29, .35)
Constant	2.122***	2.145**	1.536
	(.94, 3.30)	(.60, 3.69)	(-.10, 3.17)
Brand [€] ?	Yes	Yes	Yes
Year of sale [€] ?	Yes	Yes	Yes
Type of whiskey [€] and Region [€] ?	Yes	Yes	Yes
Distillery status [€] ?	No	Yes	Yes
Warning labels [€] ?	No	Yes	Yes
Award count?	No	Yes	Yes
Description length?	No	Yes	Yes
Bottle number and Lot no band [€]	No	Yes	Yes
Year of manufacture [€] ?	No	No	Yes
Observations	50,916	50,916	50,916
Adjusted R ²	.695	.733	.740
F Statistic	82.557***	46.13***	47.17***
DF	1421; 49494	3095; 47820	3132; 47783

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$. Model 1 is the baseline model, incorporating the basic whiskey and transaction characteristics. Model 2 extends Model 1 by including nuanced variables such as ‘Rarity,’ which indicates whether the whiskey description includes terms like ‘very rare’ or ‘rare’, and ‘Distillery status,’ identifying whether the distillery is operational, demolished, or closed; the reference level for ‘Rarity’ is the absence of rarity descriptors. Model 3, the most comprehensive model with the highest adjusted R-squared, incorporates the fixed effect of the year of manufacture. Caution is advised when interpreting the whiskey age coefficient in Model 3 due to multicollinearity with the year of manufacture.

Results

To examine the impact of time framing (year vs. length) on whiskey sold prices, we employed a linear regression model using the ‘lm’ function in R. The model regressed the log transformed sold price against the binary-coded time framing variable (1 for year, 0 for length). We included fixed effects for a comprehensive set of whiskey characteristics to control for potential confounders. These characteristics are: brand name, age, year of manufacture, type, region, distillery status, bottle size (in cl), delivery weight (in kg), alcohol strength (% ABV), and awards (quantified by the mention of award-related terms). Additionally, we accounted for the rarity of the whiskey, categorized as “none,” “rare,” and “very rare,” based on the descriptors used in the text. The number of characters in the whiskey description was also included to capture the extent of information provided about each whiskey. Warning labels were quantified by the character count and analyzed at three levels. We also controlled for the production details, including individual bottle number, production volume, and lot number—grouped into bands of 100 to associate whiskeys from similar distilleries. Transaction-specific fixed effects included VAT and sale year.

The results are detailed in Table 2.3. The first column shows estimates from a baseline model, including predictors such as time framing, brand name, whiskey age, bottle size, alcohol strength, delivery weight, type of whiskey, region, VAT applicability, and year of sale. The second column adds all other variables except for the year of manufacture. The third column presents a comprehensive model that includes every available variable, including the year of manufacture. It is important to note that the whiskey age coefficient in this fully controlled model should be interpreted with caution due to multicollinearity with the year of manufacture.

In Model 1, we observed a significant effect of time framing consistent with our

hypothesis: whiskeys described by their length sold for 8.9% more than those described by year ($B = -.086$, $SE = .01$, $t = -9.40$, $p < .0001$, $\eta^2_p = .002$). To put this effect into context, we compared it to the influence of VAT applicability¹², which was found to reduce the sold price by a similar magnitude of 8.9% (standardized $B = -.085$, $SE = .01$, $t = -8.96$, $p < .0001$, $\eta^2_p = .002$). This comparison highlights that the impact of time framing is as significant as the effect of VAT, with VAT leading to a decrease in bids by 8.9%, while time framing (length vs. year) results in an equivalent increase.

Additionally, we quantified the impact of time framing in terms of whiskey age. The model coefficients suggest that the effect of time framing on sold price is equivalent to a perceived age difference of 0.91 years, with whiskeys framed by year perceived as 0.91 years younger than those framed by length.

Models 2 and 3 reinforce the robustness of the time framing effect. Despite the addition of more variables in these models, the effect of time framing remains consistent. Specifically, in Model 3—the most comprehensive model—the effect persists ($B = -.081$, $SE = .01$, $t = -8.98$, $p < .0001$, $\eta^2_p = .002$), corresponding to an 8.4% difference in the sold price. Moreover, Model 3 explains 75.6% of the variation in the whiskey auction price. Considering the absence of buyer characteristic information, this R^2 value indicates a strong explanatory power of the model.

Discussion

Our findings indicate that whiskeys marketed by their length, rather than by year, are attributed higher market values, likely because they are perceived as older, thus supporting H1. We recognize the potential for endogeneity in our analysis (e.g., ease of finding the bottling year

¹² The coefficients for VAT applicability in Models 2 and 3 differ significantly from those in Model 1, likely due to multicollinearity with two variables—Rarity descriptor and Award count.

information in year framing, length framing being more common and hence norm-consistent), which suggests that these results should be approached with caution and not regarded as definitive or causal.

To mitigate these concerns and strengthen the credibility of our findings, we performed additional robustness checks, detailed in the Chapter 2 Appendix. These checks help ensure that our results are not unduly influenced by the aforementioned confounding factors. For instance, it is plausible that brands adhering to industry norms of using length framing may have superior marketing strategies or higher quality products, whereas norm-inconsistent brands, using year framing, might be perceived as offering lower quality. To address this, we focused our analysis on brands with a threshold level of heterogeneity in their use of year and length framing across products. We identified a subset of 370 brands and 11,029 auctions where about 50% of auctions used year framing, ensuring a balanced analysis and further validating our findings. The year-length effect holds in this sample, as shown in robustness check 3 in the Chapter 2 Appendix.

Moreover, to directly assess the causal effects of time framing on product valuation, we conducted subsequent studies in controlled laboratory settings. These controlled studies allow us to isolate the effect of time framing from other variables, providing stronger evidence for the impact of time framing on consumer perceptions and product valuations.

STUDY 2: USED GOODS

How does the format of time influence consumers' valuation of used goods? In Study 2, we explore this question by examining the impact of year versus length framing on participants' quality judgments and their willingness to pay (WTP) for secondhand products. Participants were shown various used furniture items, with their ages communicated either using year

framing (e.g., “bought in 2018”) or length framing (e.g., “bought 6 years ago”). Subsequently, they indicated their WTP for these items and assessed their quality. This study was pre-registered on AsPredicted (https://aspredicted.org/N4V_CP4).

Figure 2.3: Sofa Descriptions in Study 2

Year Framing	Length Framing
Beige Fabric Sofa, Bought in 2022  Bought in 2022; well-maintained. Material: High-quality fabric in light beige, comfortable and durable. Condition: Shows minor wear with slight creases and a bit of fading, adding character Design: Classic silhouette with clean lines, fits various decor styles. How much would you be willing to pay for this sofa (\$20 to \$1000)?	Beige Fabric Sofa, Bought 2 years ago  Bought 2 years ago; well-maintained. Material: High-quality fabric in light beige, comfortable and durable. Condition: Shows minor wear with slight creases and a bit of fading, adding character Design: Classic silhouette with clean lines, fits various decor styles. How much would you be willing to pay for this sofa (\$20 to \$1000)?

Participants and Procedure

In March 2024, we recruited 399 participants from the Prolific panel of US participants, who received monetary compensation for their participation. Based on our pre-registered exclusion criteria, we removed participants under the following conditions: duplicate IP addresses ($n = 2$), failure of the attention check ($n = 2$), and response times that were three standard deviations above or below the mean ($n = 5$). This left us with 1170 product evaluation responses from 390 participants (44.9% female, 53.8% male, 1.3% non-binary, $M_{\text{age}} = 41.5$ years). We further excluded 21 responses with WTP values more than three standard deviations above or below the mean, resulting in 1149 responses used for analysis.

Participants were randomly assigned to either the year or length time framing condition (see sample stimuli in Figure 2.3). Each participant evaluated three used furniture items (a three-seater sofa, an office chair, and a dining table set) listed on Craigslist and available in their local area, along with age information. Each item had a different age (2, 4, and 6 years) and was presented in random order, one at a time. In the year framing condition, the age of the product was described with the specific purchase year (e.g., “Bought in 2018”), whereas in the length condition, it was described with the number of years ago it was purchased (e.g., “Bought 6 years ago”). Participants indicated their WTP for each product (between \$20 and \$1000) using a text box. After indicating their WTP, participants rated the quality of each product using a five-point scale (1 Star to 5 Stars). At the end of the study, participants answered an attention check question and provided demographic information such as age, gender, income, and education.

Analyses and Results

To investigate the impact of time framing on participants’ willingness to pay (WTP) and quality evaluations, we conducted two 2 x 3 mixed factorial ANOVAs. In each ANOVA, time framing (year vs. length) was the between-subjects factor, and product type (sofa, chair, dining table) was the within-subjects factor. We log-transformed participants’ WTP data for the first ANOVA and used participants’ quality judgments for the second ANOVA.

Table 2.4: Willingness to Pay and Quality Judgments in Study 2

Product [Age]	Willingness to Pay (\$)		Quality Judgments	
	Year	Length	Year	Length
Sofa [2 years]	\$236.4	\$169.1	3.42	3.32
Office Chair [4 years]	\$82.6	\$63.1	3.16	2.90
Dining Table [6 years]	\$230.3	\$183.5	4.02	3.83
<i>Average</i>	<i>\$182.2</i>	<i>\$138.0</i>	<i>3.53</i>	<i>3.35</i>

Table 2.4—This table shows summary of results for Study 2. The values reported are raw means.

The results showed that participants were, on average, willing to pay 32% more (28% for log-transformed distributions) for products described using year framing compared to length framing ($M_{\text{Year}} = 4.77$, $SD = 1.01$; $M_{\text{Length}} = 4.52$, $SD = .94$; $F(1, 384) = 13.41$, $p = .0003$, $\eta^2_p = .03$). Additionally, participants in the year framing condition rated products as higher quality than those in the length framing condition ($M_{\text{Year}} = 3.53$, $SD = .94$; $M_{\text{Length}} = 3.35$, $SD = .97$; $F(1, 384) = 9.76$, $p = .002$, $\eta^2_p = .02$). See Table 2.4 for details. The interaction between time framing and product type was not significant in either ANOVA (log WTP: $F(2, 755) = 1.69$, $p = .19$, $\eta^2_p = .00$; Quality: $F(2, 755) = 1.72$, $p = .18$, $\eta^2_p = .00$).

Mediation Analysis. We aggregated individual participants' log WTP and quality judgment ratings by calculating the mean across the three responses provided by each participant. We then implemented a mediation analysis (Hayes, 2017, Model 4) with the time framing serving as the independent variable, the averaged log WTP as the dependent variable, and the mean quality evaluation as the mediating variable. Quality evaluation mediated the relationship between the time framing and log WTP (standardized indirect effect: $B = .12$; 95% CI = [.02, .22]). See Chapter 2 Appendix for a visual representation of the mediation model.

Discussion

Study 2 demonstrates that the representation format of the temporal attribute—year versus length—significantly impacts consumer judgments and purchase decisions. Specifically, year framing, as opposed to length framing, positively influences product quality inferences for used goods, leading consumers to be willing to pay higher prices, thus supporting H2.

While Study 1 showed that length-framed whiskeys garnered higher prices, Study 2 revealed that year-framed used goods fetched higher prices. These findings align when considering consumer preferences: older is better for whiskey, while newer is better for

secondhand furniture. Thus, the framing effect corresponds with consumers' temporal preferences for different product categories.

STUDY 3A: PAST FINANCIAL INVESTMENTS

To test the generalizability of the year length framing effect beyond the domains of liquor and used goods, we examined its influence in the context of financial investments. We employed a mixed 2 (time framing: year vs. length) x 4 (investment scenarios: two real-estate and two stock investments) design. Time framing served as a between-subjects factor, while investment scenarios acted as a within-subjects factor. Participants were randomly assigned to either the year or length framing condition and were asked to evaluate the perceived quality of investment returns for four scenarios. Each scenario included four key elements: investment type (e.g., Apple stock), amount invested (e.g., \$1k), present value (e.g., \$30k), and investment time (e.g., in 2007 or 15 years ago). The investment details were consistent across participants, with the only variation being the time framing format—year or length. See Table 2.5 for a detailed breakdown of the scenarios.

We hypothesized that time distances presented in the year framing would be perceived as shorter, leading to more favorable investment return ratings compared to the length framing. This study was pre-registered on AsPredicted (https://aspredicted.org/blind.php?x=BTY_1JF).

Method

Participants. In June 2022, we recruited 450 participants using the CloudResearch MTurk Toolkit in exchange for monetary compensation. Per our pre-registered exclusion criteria, we removed any participant whose total response time for the four dependent measures ($n = 5$) or response time to complete the survey ($n = 10$) was three standard deviations above or below the

mean response times. In total, we excluded 14 participants, leaving 436 participants for the analysis (48.9% female, 50.7% male, .5% non-binary, $M_{\text{age}} = 41$).

Procedure. Participants were randomly allocated to one of two conditions with differing time framings (year vs. length). In both conditions, they were asked to evaluate returns on investments in four scenarios, presented in random order, comprising two stocks and two real-estate investments. For instance, in a stock investment scenario, participants imagined investing \$1k in Apple stock in 2007 [15 years ago], with a current value of \$30k. They then provided their subjective investment return evaluation. Those in the year framing condition saw the year format (i.e., 2007), while those in the length framing condition received the same information but with the length format (i.e., 15 years ago). After evaluating all four investments, participants answered demographic questions and indicated whether they invest in stocks or own a house.

Dependent Variable. For each investment scenario, participants rated the perceived quality of the investment return on a nine-point scale ranging from “not a great return” to “extremely good return.” This subjective evaluation of return on investment served as the dependent variable.

Results

Subjective Investment Return. For each investment scenario, subjective investment return evaluations were measured individually, enabling a 2 x 4 analysis. These evaluations were submitted to a 2 x 4 mixed factorial ANOVA with time framing condition (year vs. length) as a between-subjects factor and investment scenarios (four types) as a within-subjects factor. The analysis revealed a significant main effect for time framing ($M_{\text{Year}} = 8.09$, $SD = 1.24$ vs. $M_{\text{Length}} = 7.82$, $SD = 1.58$; $F(1, 434) = 8.08$, $p = .005$, $\eta^2_p = .02$) and for investment scenario ($F(3, 1302) = 122.83$, $p < .0001$, $\eta^2_p = .22$). Importantly, the two-way interaction was not significant ($F(3,$

1302) = .80, $p = .49$, $\eta^2_p = .00$); however, the significance level of the time framing effect varied across the four scenarios [AMZN (7 years): $F(1,1122) = 1.38$, $p = .24$; AAPL (15 years): $F(1,1122) = 5.70$, $p = .017$; NYC (20 years): $F(1,1122) = 5.20$, $p = .023$; LA (32 years): $F(1,1122) = 7.53$, $p = .006$], such that the effect was stronger for longer time periods. Table 2.5 provides descriptive statistics by time framing and investment scenario.

Table 2.5: Past and Future Investments in Studies 3A and 3B

Framing	Time	Return Evaluation	Investment	Return	Asset
Year	2015	7.23	\$1k	\$8k	Amazon stock (AMZN)
Length	7 years	7.08	\$1k	\$8k	Amazon stock (AMZN)
Year	2007	8.28	\$1k	\$30k	Apple stock (AAPL)
Length	15 years	7.97	\$1k	\$30k	Apple stock (AAPL)
Year	2002	8.45	\$10k	\$300k	Condo in NYC
Length	20 years	8.16	\$10k	\$300k	Condo in NYC
Year	1990	8.42	\$10k	\$400k	Beach House in LA
Length	32 years	8.07	\$10k	\$400k	Beach House in LA

Framing	Time	Return Evaluation	Investment	Return	CD Institution
Year	2025	5.59	2k	\$4k	US Credit Union
Length	3 years	5.13	2k	\$4k	US Credit Union
Year	2027	7.64	\$10k	\$25k	National US Bank
Length	5 years	7.56	\$10k	\$25k	National US Bank
Year	2030	7.50	\$5k	\$15k	International Fund
Length	8 years	6.86	\$5k	\$15k	International Fund
Year	2037	7.13	\$4k	\$25k	International Bank
Length	15 years	6.71	\$4k	\$25k	International Bank

NOTE.—Study 3A (top panel) and Study 3B (bottom panel) were conducted at about the halfway point of 2022 (i.e., between the last week of June and the first week of July).

Discussion

Study 3A shows that investments described using year framing are evaluated more favorably than those using length framing, likely due to shorter perceived intervals. This suggests that the year-length effect extends to financial investments and decisions.

The findings from Studies 1, 2, and 3A consistently suggest that year framing is perceived as closer for past time intervals. One reason might be that boundary years (e.g., 2012) are more likely associated with memorable events, such as buying a car or graduating from college, unlike length framings (e.g., 12 years). This aligns with research on how the memorability or emotionality of a time can influence subjective assessments of elapsed time (Friedman 1993; Morwitz 1997; Zauberger et al. 2010; Zauberger, Ratner, and Kim 2009). To further understand the underlying mechanism and rule out memorability as an explanation, we next test whether year-length effect applies to future time distances as well. If the effect persists for future intervals, explanations related to the memorability or emotionality of past events can be dismissed. The next study is designed to explore these contrasting predictions and test the generalizability of the year-length effect.

STUDY 3B: FUTURE FINANCIAL INVESTMENTS

Study 3B aimed to examine whether the year-length effect also applies to future time periods. This study employed a design similar to Study 3A with modifications to the stimuli to focus on future time distances. Instead of evaluating past investment returns, participants assessed future returns from certificate of deposit (CD) products. CDs are fundamental financial instruments studied in prior research (LeBoeuf 2006; Monga and Bagchi 2012). They involve depositing an initial sum for a fixed term, and at the end of the maturity period, a higher amount is returned to the depositor. As the perceived maturity period lengthens while the initial deposit and return remain constant, the CD's relative attractiveness decreases.

Study 3B utilized a mixed 2 (time framing: year vs. length) x 4 (CD products) design, with time framing as a between-subjects factor and CD products as a within-subjects factor.

Participants were assigned to either the year or length framing condition. They were asked to evaluate the perceived quality of investment returns for each of the four CD products. Each CD product included four components: the institution offering the CD, investment amount, return amount at maturity, and maturity time. Across participants, all CD details remained the same except for the framing used to describe the CD's maturity time. Half of the participants saw the maturity time in the year format (e.g., in the year 2025), while the other half saw it in the length format (e.g., 3 years).

In line with our theory, we hypothesized that the year-length effect is not exclusive to past time distances. Consequently, we predicted that CDs would be rated more favorably in the year (vs. length) framing condition, as the return would be perceived as accruing over a shorter time period. The study was pre-registered on AsPredicted (aspredicted.org/blind.php?x=JS9_ZVY).

Method

Participants. In July 2022, we recruited 450 participants using the CloudResearch MTurk Toolkit in exchange for monetary compensation. Per our pre-registered exclusion criteria, we removed any participant who met any of the following conditions: duplicate IP addresses ($n = 2$); total response time for the four dependent measures ($n = 5$); or response time to complete the survey ($n = 12$) was three standard deviations above or below the mean response times. In total, we excluded 17 participants, leaving 433 participants for the analysis (45.7% female, 53.8% male, .5% non-binary, $M_{\text{age}} = 40$).

Procedure. Participants were randomly assigned to one of two conditions with different time framing (year vs. length). In both conditions, they were asked to evaluate the return on investments of four CD products, shown in random order. Before presenting the CD products,

we provided a brief explanation of the basic details of certificates of deposit to all participants. Participants then indicated their subjective investment return evaluations for the CD products. Those in the year framing condition saw the CDs' maturity times described using the year format (e.g., 2025), while those in the length framing condition saw the same CD information with maturity times described using the length format (e.g., 3 years). After evaluating all four CDs, participants answered standard demographic questions and indicated whether they had invested in a CD before.

Dependent Variable. For each CD product, participants rated their evaluation of the investment return on a nine-point scale, ranging from “not a great return” to “extremely good return.” This subjective evaluation of return on investment served as the dependent variable.

Results

Subjective Investment Return. Similar to Study 3A, the subjective investment return evaluations for different CD products were submitted to a 2 x 4 mixed factorial ANOVA with time framing condition (year vs. length) as a between-subjects factor and CD products (four types) as a within-subjects factor. The analyses revealed a significant main effect of the time framing ($M_{\text{Year}} = 6.96$, $SD = 1.85$ vs. $M_{\text{Length}} = 6.57$, $SD = 2.20$; $F(1, 431) = 8.46$, $p = .004$, $\eta^2_p = .02$) and a main effect of the CD product ($F(3, 1293) = 221.34$, $p < .0001$, $\eta^2_p = .34$). Furthermore, the two-way interaction was significant ($F(3, 1293) = 3.17$, $p = .02$, $\eta^2_p = .01$); the significance level of the time framing effect varied across the four scenarios [3 years: $F(1, 1069) = 6.59$, $p = .010$; 5 years: $F(1, 1069) = .182$, $p = .67$; 8 years: $F(1, 1069) = 12.87$, $p < 0.001$; 15 years: $F(1, 1069) = 5.50$, $p = .019$], with the effect being strongest for the 8-year CD and weakest for the 5-year CD. See Table 2.5 for descriptive statistics summarized by each factor.

Discussion

Studies 3A and 3B establish that the year-length effect influences subjective time perception and occurs for both past and future time periods, thus supporting H3. Furthermore, the data reveal that the effect size of the time framing effect is approximately the same for future ($\eta^2_p = .02$) and past ($\eta^2_p = .02$) scenarios, even though such a comparison may not be entirely accurate, as the stimuli and time distances used in Studies 3A and 3B were different.

Thus far, our first four studies demonstrate that the year-length effect is causal, manifests in at least three different domains, and holds for both past and future time periods. It is important to note that several mechanisms listed in Table 2.1 (e.g., mechanisms 1, 4, 5, 7, and 8) cannot explain these findings, as they are not applicable to the context of product evaluation. While our initial studies help rule out many mechanisms, they do not provide in-depth theoretical insights. Consequently, the following studies aimed to explore the underlying mechanism and present process evidence for the year-length effect.

STUDY 4: SUBJECTIVE TIME PERCEPTION

In the previous studies, we demonstrated that time framing could impact people's valuations and judgments in contexts where the length of time is critical, such as whiskey, used goods, and investment return evaluations. These findings suggest that time framing alters people's subjective perception of time length. However, we have not yet furnished conclusive evidence to substantiate this claim. Consequently, Study 4 had two primary aims.

First, we aimed to directly test the hypothesis that time framing alters people's time perceptions. Second, we sought to establish the generalizability and domain independence of the time framing effect. To achieve these first two aims, we asked participants to indicate how

distant or recent various time points feel relative to the present.

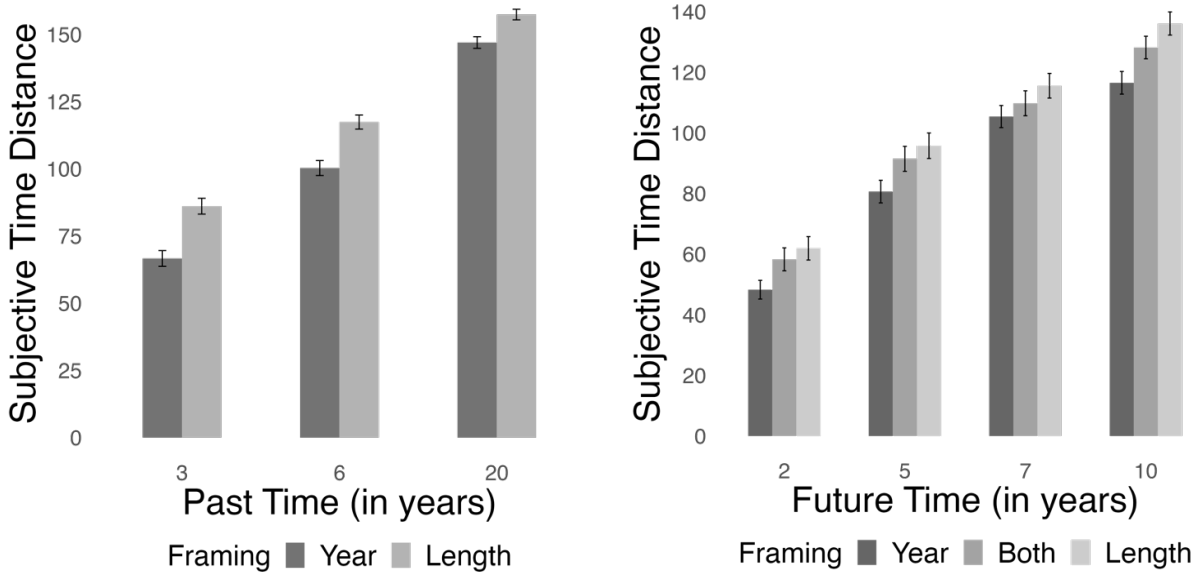
Study 4 employed a mixed 2 (time framing: year vs. length) x 3 (time distance: 3, 6, and 20 years ago) design, with time framing as a between-subjects factor and time distance as a within-subjects factor. Participants were assigned to either the year or length framing condition and asked to indicate how distant or recent each of the three past time points felt to them from the present. According to our theory, time distances are perceived as shorter when framed by the year (vs. length). The study was pre-registered on AsPredicted (aspredicted.org/TD1_HGX).

Method

Participants. In exchange for monetary compensation, we recruited 650 participants in March 2022 using the CloudResearch MTurk Toolkit. Per our pre-registered exclusion criteria, we removed participants who met any of the following conditions: duplicate IP addresses ($n = 10$); total response time for the three dependent measures ($n = 8$), or response time to complete the survey ($n = 13$) was three standard deviations above or below the mean response times. In total, we excluded 30 participants, leaving 620 participants (51.6% female, 47.9% male, .5% non-binary, $M_{\text{age}} = 41$).

Procedure. Participants were randomly assigned to one of two conditions, each featuring a different time framing (year vs. length). In both conditions, participants were presented with three randomly ordered past times and asked to rate how distant or recent each time felt to them, which served as the primary dependent variable. For example, “How far does the year 2002 [20 years ago] feel?” Participants used a non-numerical slider-scale anchored at ‘feels very recent’ (number 0) and ‘feels very distant’ (number 180) to indicate their perception of time, with only the anchors visible and the numbers hidden. Upon completing the time perception ratings for all three past times, participants answered standard demographic questions.

Figure 2.4: Subjective Time Perception in Study 4 and in the Alternate Version of Study 4



NOTE.— The left panel displays results from Study 4, while the right panel presents results from the alternate version of Study 4.

Results

Subjective Time Perception. The subjective time perception responses were submitted to a 2 x 3 mixed factorial ANOVA with time framing condition (year vs. length) as a between-subjects factor and time distance (3, 6, and 20 years) as a within-subjects factor. The analyses revealed a significant main effect of the time framing ($M_{\text{Year}} = 104.62$, $SD = 57.07$ vs. $M_{\text{Length}} = 120.27$, $SD = 53.47$; $F(1, 618) = 26.14$, $p < .0001$, $\eta^2_p = .04$) and a main effect of the time distance ($F(2, 1236) = 932.19$, $p < .0001$, $\eta^2_p = .60$). See left panel of Figure 2.4. Moreover, the two way interaction was significant ($F(2, 1236) = 3.55$, $p = .03$, $\eta^2_p = .006$), such that the time framing effect was stronger for 3 and 6 years compared to 20 years; see Chapter 2 Appendix for means summarized by time framing and time distances.

Discussion

Study 4 provides process evidence for the year-length effect. The findings indicate that time framing (year vs. length) alters people's subjective perceptions of time, making year-framed

times feel less distant compared to length-framed times.

Alternate version of Study 4. To explore the impact of using both year and length formats simultaneously when describing time intervals, we conducted a modified version of Study 4 in January 2023 ($n = 408$, 46% non-male, $M_{\text{age}} = 40$). We made two changes to the study design: First, we added a third condition in which time intervals were described using both year and length formats. For instance, to describe a time interval two years in the future, participants saw “same time of the year 2025” on the first line and “2 years later” on the second line, with the descriptions vertically stacked. See Chapter 2 Appendix for the exact stimuli. Participants then indicated how distant this time interval felt. Second, we focused on future time intervals to establish the generalizability of the effect of time framing on subjective time perception. This study employed a mixed 3 x 4 design, with time framing (year, length, both year and length) as a between-subjects factor and time distance (2, 5, 7, and 10 years later) as a within-subjects factor. The study was pre-registered on AsPredicted (aspredicted.org/K26_6NP).

We analyzed the data using a method similar to Study 4. The main effect of time framing was significant ($F(2, 405) = 4.34, p = .01, \eta^2_p = .02$), with subjective distances following a descending order from length, dual, to year framing. The year framing was significantly different from length framing ($F(1, 405) = 8.48, p = .004, \eta^2_p = .02$), and dual framing marginally different from year framing ($F(1, 405) = 3.35, p = .07, \eta^2_p = .01$). However, no significant difference was found between the length and dual framing conditions ($F(1, 405) = 1.15, p = .28, \eta^2_p = .00$). The interaction between time framing and time distance was marginally significant ($F(6, 1215) = 1.83, p = .09, \eta^2_p = .01$). The right panel of Figure 2.4 visually presents these results. See Chapter 2 Appendix for detailed means.

Participants’ average time distance judgments in the dual framing condition fell between the means of the year and length framing conditions, suggesting that both year and length descriptions

play a role in shaping people's judgments in this framing condition. However, the impact of the length framing is more dominant than the year framing, indicating that people might be more influenced by the length framing than by the year framing when both formats are presented together.

STUDY 5: MEDIATION BY TIME PERCEPTION

The previous study revealed that time framing influences people's subjective assessments of time distance. Building on this, Study 4 aims to demonstrate how these subjective assessments of time distance explain the effect of time framing on people's judgments and decisions. To capture a broad range of time-related decisions across various domains and contexts within a single task, we evaluated how people discount the value of future outcomes compared to present ones. Temporal discounting is a widely used psychological tool that measures individuals' general propensity to delay gratification, and it is shown to predict many important real-world behaviors, including credit card debt, smoking, exercise, body-mass index, and infidelity (Hardisty et al. 2013). Consequently, Study 5 aims to test whether time framing-induced changes in subjective assessments of time distance can impact people's temporal discounting patterns.

This study employed a 2 (time framing: year vs. length) x 3 (time distance: 2, 3, and 4 years) design, where both time framing and time distance were between-subjects factors. Participants were assigned to one of the six cells. We first administered a fixed sequence titration task (Hardisty et al. 2013), where the future time (2, 3, or 4 years) was described using either year or length framing. We then asked participants to indicate how distant the future time felt, similar to Study 4. According to our theory, time distances are perceived as less distant when described using the year (vs. length) framing, which should result in lower discount rates for participants. The study was pre-registered on AsPredicted (aspredicted.org/blind.php?x=PQG_3BK).

Method

Participants. We recruited 2400 participants in May 2022 using CloudResearch MTurk Toolkit and compensated them monetarily. Applying pre-registered exclusion criteria, we removed participants for: duplicate IP addresses ($n = 39$); attention check failure ($n = 58$); inconsistent responses, i.e., switched back and forth, in the intertemporal choice task ($n = 31$); total response time to complete the intertemporal choice task was three standard deviations above the mean response times or less than ten seconds ($n = 167$); response time to complete the survey ($n = 34$) or to complete the subjective time perception question was three standard deviations above or below the mean response times ($n = 20$); and always choosing ‘money now’ option ($n = 375$). In total¹³, we excluded 711 participants, leaving 1689 participants for the analysis (53.0% female, 46.2% male, .7% non-binary, $M_{\text{age}} = 42$).

Procedure. Participants were randomly assigned to one of six cells, with variations in time framing (year vs. length) and time distance (2, 3, or 4 years). In each condition, participants were informed that their city was offering a \$1000 rebate, but they had the option to wait and receive a larger amount in the future. We employed a fixed-sequence titration task adapted from prior research (Hardisty et al. 2013). Participants indicated the minimum amount that would justify delaying the rebate (see Chapter 2 Appendix for the exact scenario). They were presented with seven pairs of choices, comparing \$1000 now to a progressively larger amount later (e.g., \$1000 today vs. \$1500 in 2025 [in 3 years] on the same day), with the later amount increasing in each subsequent pair. To maintain similar hyperbolic discount rates, the larger later amounts for each of the seven pairs were adjusted across the three time distance conditions (see Chapter 2 Appendix for all monetary values used).

¹³ In time discounting tasks, it is common to exclude many participants, such as 39% excluded in Study 1 by Hardisty and colleagues (2013).

After completing the intertemporal choice task, participants rated their subjective time perception for the future time using a procedure similar to Study 4. The future time (2, 3, or 4 years) and the time framing (year or length) in the subjective time perception question matched the intertemporal choice task. Finally, participants answered standard demographic questions and an attention check question.

Dependent Variable. First, we determined each participant's switching point in the intertemporal choice task (i.e., when they shifted from preferring 'money now' to 'money later') to calculate their indifference point. Next, we used this indifference point to compute their discount rate for the future, applying a hyperbolic discounting function (Mazur 1987). The participant's hyperbolic discount rate served as the primary dependent variable.

Results

Discount Rates. The hyperbolic discount rates of participants were analyzed using a two-way ANOVA with time framing and time distance as between-subjects factors. The results demonstrated that participants' discount rates were lower in the year framing ($M_{\text{Year}} = .23$, $SD = .14$) than in the length framing ($M_{\text{Length}} = .25$, $SD = .13$; $F(1, 1683) = 5.83$, $p = .016$, $\eta^2_p = .003$). Additionally, a main effect of time distance was observed ($F(2, 1683) = 5.79$, $p < .01$, $\eta^2_p = .007$), indicating the highest discount rate in the two-year condition ($M_{2_years} = .25$, $M_{3_years} = .23$, and $M_{4_years} = .23$). However, the two-way interaction between the factors was not significant ($F(2, 1683) = 1.31$, $p = .27$, $\eta^2_p = .002$).

Subjective Time Perception. Subjective time perception responses were analyzed using an ANOVA with time framing condition and time distance as between-subjects factors. The analysis revealed a significant main effect of time framing ($M_{\text{Year}} = 107.11$, $SD = 45.51$; $M_{\text{Length}} = 116.44$, $SD = 43.34$ vs.; $F(1, 1683) = 17.84$, $p < .0001$, $\eta^2_p = .01$) and a main effect of the time distance ($M_{2_years} = 103.11$, $M_{3_years} = 110.41$, $M_{4_years} = 121.94$; $F(2, 1683) = 26.18$, $p <$

.0001, $\eta^2_p = .03$). Furthermore, the two-way interaction was marginally significant ($F(2, 1683) = 2.43, p = .09, \eta^2_p = .01$), with the time framing effect being stronger for 2 and 3 than for 4 years.

Mediation Analysis. We conducted a mediation analysis (Model 4, Hayes 2017) to examine our proposed conceptual framework, which suggests that the effect of time framing on discount rate is mediated by the subjective time perception. Time framing served as the independent variable, discount rate as the dependent variable, and subjective time perception as the mediator. The results showed that the indirect path was significant (standardized indirect effect: $B = -.05$; 95% CI = $[-.08, -.03]$), while the direct path was not (standardized direct effect: $B = -.06$; 95% CI = $[-.16, -.03]$). See Chapter 2 Appendix for means and a graphical representation of the mediation model.

Discussion

Study 5 offers substantial support for our conceptual model of the year-length effect. While our prior studies focused on how the time perception differences in year-length effects can affect product valuation, this study uses a more traditional dependent measure of time discounting to show that time perception also affects these types of judgments. The results demonstrate that time framing (year vs. length) has a significant impact on individuals' perception of time duration, affecting their discount rates. These findings suggest that time framing can serve as a valuable instrument for marketers and policymakers who aim to shape people's intertemporal choices involving the tradeoffs between time and monetary preferences.

STUDY 6A: PROCESS EVIDENCE THROUGH MEDIATION

Studies 4 and 5 demonstrated that the year-length framing influences the perceived magnitude of time intervals. In Study 6A, we aim to test our theory that this phenomenon is

driven by numerical change perception and test H4a. Specifically, we propose that when time intervals are framed by specific years (e.g., from 2024 to 2012), the perceived change feels smaller due to the logarithmic nature of the mental number line, where larger numbers are perceived as closer together. Conversely, when time is framed by length (e.g., 12 years ago), the change feels more substantial because each unit represents a more significant portion of the mental number line.

We employed a mixed 2 (time framing: year vs. length) x 2 (time distance: 5 and 9 years later) design, with time framing as a between-subjects factor and time distance as a within-subjects factor. Participants first assessed subjective time perception for the two time distances. Subsequently, they evaluated subjective numerical change for two pairs of numbers. In the year condition, the pairs were 1000 and 1005, and 1000 and 1009. In the length condition, the pairs were 0 and 5, and 0 and 9. We chose numbers in the thousands for the year framing and single digits for the length framing to align with our theory while ensuring the independence of the numerical perception task from the time perception task. By avoiding numbers in the 2000s, we minimized the likelihood of participants interpreting them as years.

Our hypothesis was that subjective numerical change perception would mediate the effect of year-length framing on subjective time perception. This study was pre-registered (https://aspredicted.org/D1J_PKD).

Method

Participants. In April 2024, we recruited 464 participants from the Prolific USA panel. Per our pre-registered exclusion criteria, we removed participants who met any of the following conditions: duplicate IP addresses ($n = 12$), failed the fraud detection check ($n = 10$), or had a response time to complete the survey that was three standard deviations above or below the mean

($n = 6$). In total, 28 participants were excluded, leaving 436 participants for analysis (51% female, 49% male, 0% non-binary, $M_{\text{age}} = 37$).

Procedure. Participants were assigned to either the year or the length condition and informed that the survey consisted of two sections. The first section was the subjective time perception task, similar to Study 4. In this section, participants answered two questions presented in random order using a non-numerical slider scale anchored at ‘feels very close’ (0) and ‘feels very far’ (180): “How far does the year 2029 [5 years later] feel from now?” and “How far does the year 2033 [9 years later] feel from now?” The cursor started from the point where participants clicked first on the scale, ensuring no initial bias.

In the second section, participants were informed that to understand how people perceive changes in numbers, they would be shown two numerical changes and asked how big the change felt to them. In both conditions, participants were shown two numerical changes in random order and indicated their subjective numerical change perception on a non-numerical slider scale anchored at ‘feels very small’ (0) and ‘feels very big’ (180). The two questions in the year condition were: “How big does the change from 1000 to 1005 feel to you?” and “How big does the change from 1000 to 1009 feel to you?” whereas in the length condition, the questions were: “How big does the change from 0 to 5 feel to you?” and “How big does the change from 0 to 9 feel to you?” After indicating their numerical change perception, participants answered five fraud detection check questions and then standard demographic questions. See the Chapter 2 Appendix for details about the fraud detection questions used and the exclusion criteria.

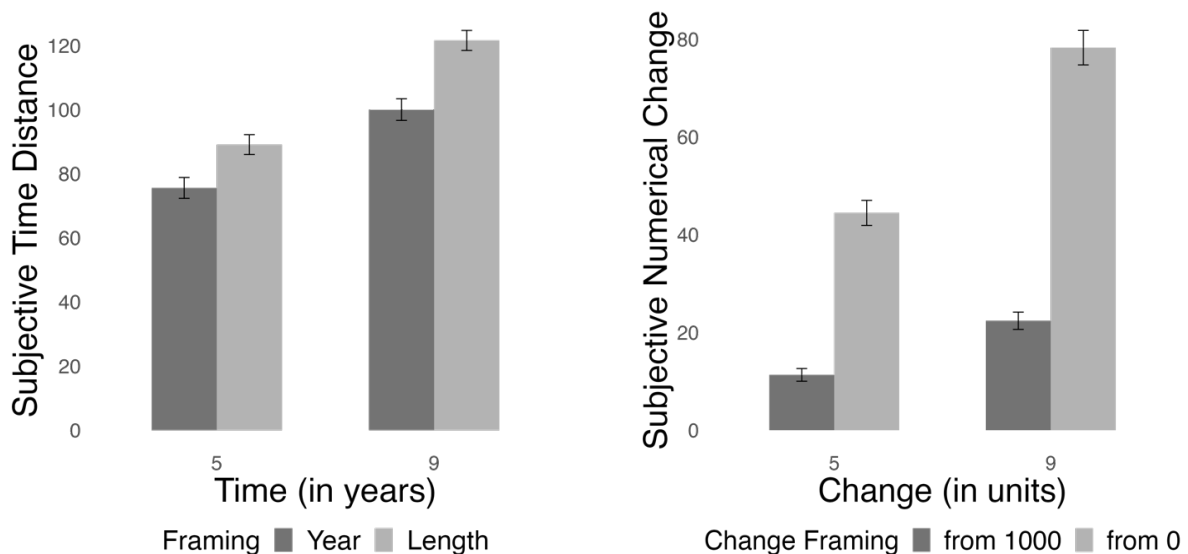
Results

Subjective Time Perception. The subjective time distance responses were analyzed using a 2 x 2 mixed factorial ANOVA, with time framing condition (year vs. length) as a between-

subjects factor and time distance (5 and 9 years) as a within-subjects factor. The analyses revealed significant main effects of time framing ($M_{\text{Year}} = 87.80$, $SD = 50.21$ vs. $M_{\text{Length}} = 105.34$, $SD = 48.85$; $F(1, 434) = 16.38$, $p = .0001$, $\eta^2_p = .04$) and time distance ($M_{5_years} = 82.34$, $SD = 47.24$ vs. $M_{9_years} = 110.91$, $SD = 49.20$; $F(1, 434) = 437.54$, $p < .0001$, $\eta^2_p = .50$). The two-way interaction between time framing and time distance was significant ($F(1, 434) = 8.76$, $p = .003$, $\eta^2_p = .02$), such that the time framing effect was stronger for 9 years than 5 years; see left panel of Figure 1.5 for a graphical representation.

Subjective Numerical Change. We submitted the subjective numerical change responses also to the same ANOVA model as above. The results were similar; the main effects of framing ($M_{\text{change_from_1000}} = 16.85$, $SD = 23.43$ vs. $M_{\text{change_from_0}} = 61.36$, $SD = 48.76$; $F(1, 434) = 180.10$, $p < .0001$, $\eta^2_p = .29$) and unit change ($M_{5_units} = 28.04$, $SD = 34.41$ vs. $M_{9_units} = 50.58$, $SD = 49.98$; $F(1, 434) = 471.66$, $p < .0001$, $\eta^2_p = .52$) were significant. The two-way interaction was also significant ($F(1, 434) = 121.01$, $p < .0001$, $\eta^2_p = .22$), such that the framing effect was stronger for 9 unit change than 5 unit change; see right panel of Figure 2.5.

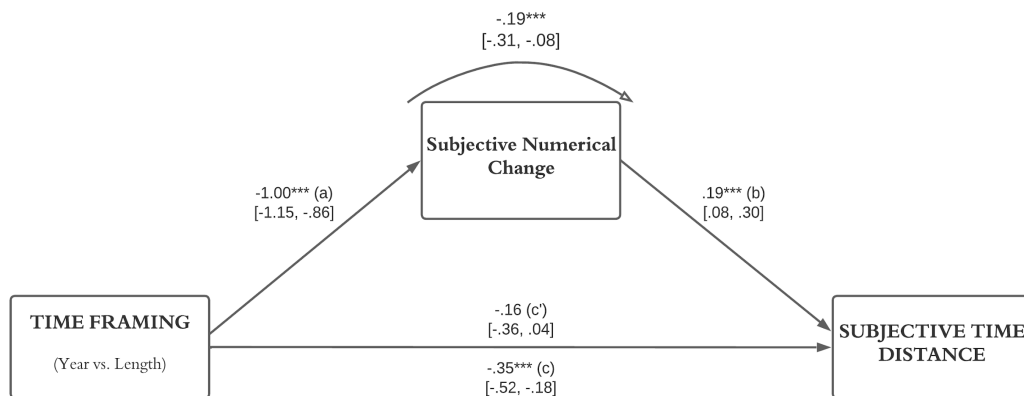
Figure 2.5: Subjective Time Distance and Subjective Numerical Change in Study 6A



NOTE.— The left panel displays subjective time distance while the right panel shows the subjective numerical change observed in Study 6A.

Mediation Analysis. To test our theory, we used a mediation model (Model 4, Hayes 2017) with a bootstrapping method to test whether the effect of time framing on subjective time distance is mediated by subjective numerical change perception. We calculated the average subjective time distance and average subjective numerical change for each participant by averaging their responses for the two time distances. Time framing was used as the independent variable, average subjective time distance as the dependent variable, and average subjective numerical change as the mediator. The analysis showed that subjective numerical change fully mediated the effect of time framing on subjective time distance ($B = -.19$; 95% CI = $[-.31, -.08]$); see Figure 2.6.

Figure 2.6: Subjective Numerical Change Explains Year-Length Effect in Study 6A



NOTE.— All continuous variables are standardized. Numbers annotating curved arrows indicate indirect effects. Dashed lines show statistically not significant paths. Bracketed numbers indicate 95% CIs. * $p < .05$, ** $p < .01$, *** $p < .001$, ^ $p < .05$, exact value N/A.

Discussion

These results provide evidence that the year-length effect on time perception is driven by the logarithmic nature of the mental number line, where equal differences between larger numbers are perceived as smaller than those between smaller numbers. For instance, larger numbers (e.g., 1000 to 1005) were perceived as closer together than smaller numbers (e.g., 0 to

5). This perception translated to time intervals, with time framed by years (e.g., 2024 to 2029) feeling shorter than the same time framed by length (e.g., 5 years later). Understanding this mechanism allows us to potentially switch off the year-length effect. Therefore, in the next study, we use this insight to design interventions aimed at attenuating the year-length effect, providing further evidence supporting our theorization.

STUDY 6B: PROCESS EVIDENCE THROUGH MODERATION

Study 6B aimed to further support our theorization by investigating the role of numerical change perception in the year-length effect. As numbers increase, the perceived difference between two consecutive numbers diminishes because of the logarithmic shape of the mental number line. In other words, smaller numbers are mapped more linearly, whereas larger numbers follow a logarithmic representation. Since the nominal values in year framing are significantly larger than those in length framing, year framed times involve logarithmic representations, while length framed times are relatively linearly mapped.

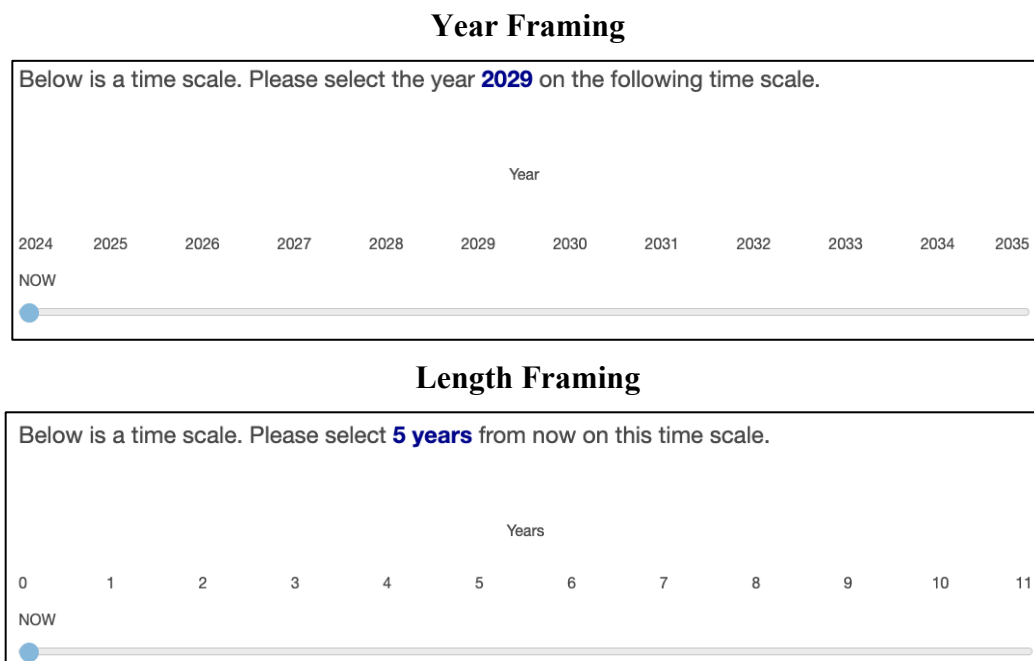
If this hypothesis is correct, then temporarily superimposing a linear scale should mitigate the year-length effect. Specifically, we hypothesized that introducing a linear scale would make participants' responses in the year framing condition resemble those in the length framing condition. This study was pre-registered (https://aspredicted.org/RQ8_LHY).

Method

Participants. In March 2024, we recruited 1,003 participants using the Prolific USA panel. Per our pre-registered exclusion criteria, we removed participants who met any of the following conditions: duplicate IP addresses ($n = 8$), failed the fraud detection check ($n = 13$), response time to complete the survey was three standard deviations above or below the mean

response times ($n = 17$), or provided incorrect answers in the linear scale superimposition task ($n = 10$). In total, we excluded 48 participants, leaving 965 participants for the analysis (50.4% male, 47.6% female, 2% non-binary, $M_{\text{age}} = 41$).

Figure 2.7: Linear Scale Superimposition in Study 6B



NOTE.— The top image shows the linear time scale superimposition in the year framing while the bottom panel shows the linear time scale superimposition in the length framing.

Procedure. We employed a mixed 2 (time framing: year vs. length) x 2 (scale: default vs. linear) x 2 (time distance: 5 and 9 years later) design, with time framing and scale as between-subjects factors and time distance as a within-subjects factor. In the default scale conditions, participants answered two questions presented in random order using a non-numerical slider scale anchored at ‘feels very close’ (0) and ‘feels very far’ (180): “How far does the year 2029 [5 years later] and the year 2033 [9 years later] feel from now?” This process mirrored the first section of Study 6A.

In the linear scale superimposition conditions, we introduced a linear scale before participants judged the subjective time perception for 2029 [5 years later] and 2033 [9 years

later]. Participants indicated the position of the future time on a linear scale. For example, they marked the position of 2029 [5 years later] before judging how far 2029 [5 years later] felt. In the year framing condition, the linear scale spanned from 2024 to 2035, with each subsequent year placed equidistantly. In the length framing condition, the scale ranged from 0 to 11 years, with each year similarly spaced (see Figure 2.7 for the linear scale superimposition). After marking the position on the linear time scale, participants answered the subjective time perception question, as in the default scale condition. Subsequently, all participants completed the fraud detection check and standard demographic questions.

Results

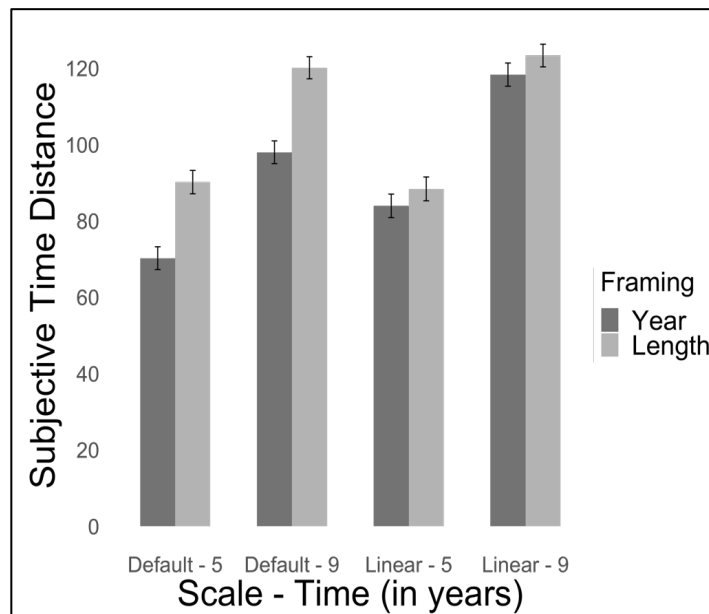
Subjective Time Perception. The subjective time distance responses were analyzed using a 2 x 2 x 2 mixed factorial ANOVA, with scale (default vs. linear) and time framing (year vs. length) as between-subjects factors and time distance (5 and 9 years) as a within-subjects factor. The analyses revealed significant main effects for time framing ($M_{\text{Year}} = 92.55$, $SD = 50.01$ vs. $M_{\text{Length}} = 105.51$, $SD = 49.09$; $F(1, 951) = 19.39$, $p < .0001$, $\eta^2_p = .02$), scale ($M_{\text{Default}} = 94.58$, $SD = 49.60$ vs. $M_{\text{Linear}} = 103.49$, $SD = 49.97$; $F(1, 951) = 9.20$, $p = .0025$, $\eta^2_p = .01$), and time distance ($M_{5_years} = 83.13$, $SD = 47.87$ vs. $M_{9_years} = 114.85$, $SD = 46.91$; $F(1, 951) = 1825.13$, $p < .0001$, $\eta^2_p = .66$).

The three-way interaction was not significant ($F(1, 951) = .31$, $p = .58$, $\eta^2_p = .00$), indicating that the two-way interactions did not vary by the third factor. Importantly, the two-way interaction between time framing and scale was significant ($F(1, 951) = 7.78$, $p = .005$, $\eta^2_p = .01$). Specifically, the time framing effect was significant in the default scale condition ($M_{\text{Year}} = 84.13$, $M_{\text{Length}} = 105.16$; $F(1, 951) = 26.17$, $p < .0001$, $\eta^2_p = .03$), but the time framing effect did not manifest when the linear scale was superimposed ($M_{\text{Year}} = 101.16$,

$M_{\text{Length}} = 105.87$; $F(1, 951) = 1.29$, $p = .26$, $\eta^2_p = .00$); see Figure 2.8 for a graphical representation. This finding is consistent with our theory, as the linear scale superimposition increased the subjective time perception for year framing to match that of length framing, and did not affect the length condition.

The two-way interaction between time framing and time distance was not significant ($F(1, 951) = .87$, $p = .35$, $\eta^2_p = .00$). Additionally, within the default scale condition, the effect of time framing did not vary with time distance ($F(1, 951) = 1.12$, $p = .29$, $\eta^2_p = .00$). However, upon removing outlier responses—values less than or equal to 2 or greater than or equal to 178 on a subjective time distance scale of 0 to 180 (about 7% of the total responses)—we observed a different pattern. In this adjusted analysis, the interaction between time framing and time distance in the default scale condition reached marginal significance. Specifically, the effect of time framing was found to be stronger for 9 years compared to 5 years ($F(1, 843) = 3.34$, $p = .067$, $\eta^2_p = .00$), mirroring the pattern observed in Study 6A.

Figure 2.8: Linear Scale Superimposition Attenuates the Year-Length Effect



NOTE.— The figure shows raw means from Study 6B and bars depict standard errors.

Discussion

Studies 6A and 6B collectively provide process evidence supporting the numerical change perception hypothesis (H4a and H4b) that underlies the year-length effect. We argued that larger nominal values in year framing follow a logarithmic representation, while smaller values in length framing are mapped more linearly, influencing time perception. Consistent with this theorization, superimposing a linear scale made the year framing condition behave similarly to the length framing condition. Assuming that linear representation of quantities is a more accurate way to represent them than a logarithmic representation, this indicates that length framing activates more accurate time perceptions, whereas year framing biases those perceptions.

GENERAL DISCUSSION

Time-based decisions are integral to our daily lives, influencing how we plan and make choices based on perceived temporal distances. Both year framing (e.g., year 2019 or year 2026) and length framing (e.g., 5 years ago or 2 years later) are common methods for communicating time intervals that span over a year. Our research explored the impact of these time interval framings on consumers' evaluative judgments and subjective time perception through a series of experiments. Study 1 validated the effect in a real-world context using archival data, showing that whiskeys were valued more when their age was marketed using length framing compared to year framing. Studies 2, 3A, and 3B examined the influence of year-length framing on willingness to pay for used products and financial product evaluations. Studies 4 and 5 demonstrated that time framing affects people's subjective time perceptions, which in turn influences their intertemporal choices. Finally, Studies 6A and 6B provided process evidence

that subjective numerical change perception drives the year-length effect.

Besides the numerical change perception explanation, several alternative mechanisms could potentially account for the year-length effect. Table 2.6 summarizes how our studies systematically rule out some of the prominent alternative explanations discussed in prior research on date-duration framing effects.

Table 2.6: Discussion of Alternate Accounts of Year-Length Effect

Alternate Account	Ruled out by	Comments
Vividness of duration	Studies 6B and Supplementary Study 1	Prior research (LeBouef, 2006) suggests that duration is more vivid in length framing (e.g., “2 years”) compared to year framing (e.g., “2026”), which could increase perceived duration and contribute to the year-length effect. We rule out this vividness account in two key ways. First, Supplementary Study 1 directly measured vividness and found that participants rated vividness significantly higher in year framing, contrary to expectations, suggesting vividness does not explain the framing effect. Second, Study 6B shows that the year-length effect disappears when a linear scale is superimposed, as the year condition aligns with the length condition under this manipulation, further indicating that vividness is not the driving factor.
Ease of computation	Studies 6B, Supplementary Study 1, and Supplementary Study 2	People need to compute the time distance in year framing to judge the subjective time distance, whereas no such computation is required in length framing. This difference in computational ease might explain the link between year-length framing and subjective time distance (Thomas and Morwitz 2009b). However, our findings from Study 6B indicate that the linear scale superimposition manipulation completely attenuates the year-length effect on subjective time distance, effectively ruling out the computational ease explanation. Additionally, in Supplementary Study 1 and 2, neither the computational ease manipulation nor the self-reported measurements of ease supported this account.
Preference /processing fluency/ familiarity	Studies 1, 2, 3A, and 3B	This account posits that one framing might be easier to process, preferred, or more familiar to consumers, leading them to favor products presented in that framing. However, the results from Studies 1, 2, 3A, and 3B demonstrate a reversal of patterns, which contradicts this explanation. Specifically, while consumers were willing to pay higher prices for length-framed whiskeys, they were willing to pay lower prices for length-framed used goods and financial products. This inconsistency across different product domains indicates that simple preference, fluency, or familiarity does not drive the year-length effect.
Domain specific norms	Studies 4, alternate version of 4, 6A, and 6B	It is possible that the year-length effect observed in our studies is driven by context or industry norms rather than being a generalizable phenomenon. For instance, in the whiskey industry, it is common to use length framing, while in the market for used goods, year framing is the norm. This could suggest that our findings are merely capturing participants’ higher willingness to pay for products that conform to industry norms. However, this explanation does not account for the results of Studies 4, 6A, and 6B, which were context-independent and simply asked participants to evaluate how far in the future or

		past a given time felt.
Precision/ specificity	Study 6B	Prior research suggests that year framing appears more precise or specific, leading participants to feel more certain about receiving future rewards. Consequently, they might be more willing to wait for future rewards when described using year framing compared to length framing. However, this explanation fails to account for the year-length framing effect in the context of product valuation and subjective time distance, as these scenarios do not involve future rewards. Additionally, Study 6B demonstrates that the year-length effect is completely attenuated by the linear scale superimposition manipulation, further ruling out the precision explanation.

Impact on Response Times

We tracked participants' response times in answering the subjective time perception question in all studies where we measured subjective time perception. Our findings reveal that response times are shorter with length framing compared to year framing, and these response times can mediate the effect of time framing on subjective time perception. This observation is consistent with our theory that time perception is influenced by number perception. In scenarios where nominal values are larger, as typically seen in year framing, the ease of distinguishing magnitudes at the same numerical distance (e.g., 5 and 15 vs. 90 and 100) decreases—a phenomenon known as the size effect (Parkman, 1971; also see problem size effect by Zbrodoff and Logan 2005). This leads to longer processing times for larger numbers due to the increased difficulty in assessing magnitude differences or computing them. However, it is important to emphasize that while response latency is concomitant with the observed effects, it does not causally explain the effect of time framing on magnitude difference judgment. In other words, the shorter response times in length framing do not directly cause the perceived differences in time but rather accompany them in a way that aligns with our theory.

Theoretical Implications

Our proposed mechanism contributes to the theoretical landscape in three significant ways. First, prior research demonstrates that consumers' perceptions of time are nonlinear,

explaining phenomena such as hyperbolic discounting, where a year may feel less than twelve times as long as a month (Zauberman et al. 2009). Our work extends this idea by showing that the nonlinearity in time perception can be further accentuated by the nonlinear nature of the internal mental number line, especially when processing large nominal values used to denote time (e.g., 2028).

Second, we demonstrate that time perception is affected by the nominal values used to denote time, indicating that it is influenced by the number perception system. While prior research has explored links between number and space, the connections between number and time or space and time have been less frequently studied, and the shared properties of all three systems have not been thoroughly considered (Walsh 2003). Our empirical evidence supports Walsh's (2003) theory that time, space, and numerical quantity share a common cognitive processing framework, suggesting these dimensions could be part of a generalized magnitude system. This integrated view implies that our brains may use a unified system to process different types of magnitudes, offering a holistic approach to understanding cognitive processing across various dimensions.

Third, because there is no definitive answer to how far a time point (e.g., 3 years later) seems from the present, previous research on time interval framing often only compares relative magnitude judgments between different framings without indicating which framing causes over- or underestimation. However, our work within the landscape of time interval framing is unique in this regard. Based on our process evidence, we can infer that length framing (e.g., 3 years) activates more accurate time perceptions, while year framing (e.g., 2027) biases those perceptions due to the logarithmic nature of the mental number line. We base this inference on the assumption that a linear representation of quantities is more accurate than a logarithmic

representation.

Managerial Implications

Our research demonstrates the importance of time framing in effective communication for both brands and consumers. The following examples illustrate how our findings can be applied to enhance communication strategies and outcomes:

Corporate Communication: Time framing can play a crucial role in corporate communication and investor relations. For example, when a CEO in an investor pitch meeting wants to signal a longer runway based on existing funds, they have two choices: “We have funds to survive for the next 2 years” or “We have funds to survive until 2026.” Using length framing (i.e., 2 years) can more effectively emphasize the longevity of the company’s runway compared to year framing (i.e., 2026). Additionally, companies often highlight the number of years they have been in operation to signal a long and rich history. While updating the number of years on a permanent facade annually may be impractical, marketing campaigns and digital press releases can be easily updated on a regular basis. See Chapter 2 Appendix for examples of companies that use length framing to signal their history.

Product Marketing: Our research offers valuable insights for industries where the temporal attribute affects product value, such as antiques, fine art, liquor, aged food products like cheese, meat, and chocolates, as well as used goods. In these contexts, using optimal framing strategies can enhance perceived value and sales, depending on whether a longer perceived time interval is beneficial (e.g., whiskey) or detrimental (e.g., used furniture) to valuation.

Welfare and Non-Profit Organizations: Time framing can also be a powerful tool for welfare organizations, such as climate-focused initiatives. By framing goals in year terms, organizations can better communicate the urgency of their mission. For example, “Join to take

action now to reach net-zero carbon by 2040” is likely to result in more people joining the pledge than “Reach net-zero carbon in 16 years,” as the goal seems closer and hence more impactful.

Broader Decision-Making Contexts: The practical relevance of time framing extends beyond business, influencing decision-making in various domains such as finance, health, and social behavior. Time framing can affect how people perceive the time-money tradeoffs, the perceived risk of a future event, and even goal setting and planning. For instance, framing retirement timing in terms of year (e.g., retire in 2048) rather than length (e.g., retire in 24 years) may prompt employees to save more for the future since the future seems closer.

Overall, understanding and applying the principles of time framing can help businesses and organizations communicate information more effectively, influence consumer behavior, and achieve their objectives more successfully, leading to greater impact in various contexts.

REFERENCES

- Barth, Hilary, Kristen La Mont, Jennifer Lipton, Stanislas Dehaene, Nancy Kanwisher, and Elizabeth Spelke (2006), “Non-symbolic arithmetic in adults and young children,” *Cognition*, 98 (3), 199–222.
- Brannon, Elizabeth M. and Herbert S. Terrace (1998), “Ordering of the Numerosities 1 to 9 by Monkeys,” *Science*, 282 (5389), 746–49.
- Burson, Katherine A., Richard P. Larrick, and John G. Lynch (2009), “Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals,” *Psychological Science*, 20 (9), 1074–78.
- Dehaene, Stanislas (1997), *The Number Sense*, Oxford Univ. Press, Oxford.
- Dehaene, Stanislas (2001), “Subtracting Pigeons: Logarithmic or Linear?,” *Psychological Science*, 12 (3), 244–46.
- Dehaene, Stanislas, Emmanuel Dupoux, and Jacques Mehler (1990), “Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison,” *Journal of experimental Psychology: Human Perception and performance*, 16 (3), 626.
- Dehaene, Stanislas, Véronique Izard, Elizabeth Spelke, and Pierre Pica (2008), “Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures,” *Science*, 320 (5880), 1217–20.
- Dehaene, Stanislas, Manuela Piazza, Philippe Pinel, and Laurent Cohen (2003), “Three Parietal Circuits for Number Processing,” *Cognitive neuropsychology*, 20, 487–506.
- Fechner, Gustav Theodor (1860), *Elemente der Psychophysik*, Breitkopf u. Härtel.
- Feigenson, Lisa, Stanislas Dehaene, and Elizabeth Spelke (2004), “Core systems of number,” *Trends in Cognitive Sciences*, 8 (7), 307–14.

- Friedman, William J. (1993), "Memory for the time of past events.," *Psychological bulletin*, 113 (1), 44.
- Halberda, Justin, Michèle M. M. Mazzocco, and Lisa Feigenson (2008), "Individual differences in non-verbal number acuity correlate with maths achievement," *Nature*, 455 (7213), 665–68.
- Hardisty, David J., Katherine F. Thompson, David H. Krantz, and Elke U. Weber (2013), "How to measure time preferences: An experimental comparison of three methods," *Judgment and Decision making*, 8 (3), 236–49.
- Hayes, Andrew F. (2017), *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*, Guilford publications.
- LeBoeuf, Robyn A. (2006), "Discount rates for time versus dates: The sensitivity of discounting to time-interval description," *Journal of Marketing Research*, 43 (1), 59–72.
- LeBoeuf, Robyn A. and Eldar Shafir (2009), "Anchoring on the "here" and "now" in time and distance judgments.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35 (1), 81.
- Leland, Jonathan W. (2002), "Similarity judgments and anomalies in intertemporal choice," *Economic Inquiry*, 40 (4), 574–81.
- Malkoc, Selin A., Gal Zauberman, and James R. Bettman (2010), "Unstuck from the concrete: Carryover effects of abstract mindsets in intertemporal preferences," *Organizational Behavior and Human Decision Processes*, 113 (2), 112–26.
- Mazur, James E. (1987), "An adjusting procedure for studying delayed reinforcement," *Quantitative analyses of behavior*, 5, 55–73.

- Monga, Ashwani and Rajesh Bagchi (2012), “Years, Months, and Days versus 1, 12, and 365: The Influence of Units versus Numbers,” *Journal of Consumer Research*, 39 (1), 185–98.
- Morwitz, Vicki G. (1997), “It seems like only yesterday: The nature and consequences of telescoping errors in marketing research,” *Journal of Consumer Psychology*, 6 (1), 1–29.
- Moyer, Robert S. and Thomas K. Landauer (1967), “Time required for judgements of numerical inequality,” *Nature*, 215 (5109), 1519–20.
- Munichor, Nira and Robyn A. LeBoeuf (2018), “The influence of time-interval descriptions on goal-pursuit decisions,” *Journal of Marketing Research*, 55 (2), 291–303.
- Pandelaere, Mario, Barbara Briers, and Christophe Lembregts (2011), “How to Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons,” *Journal of Consumer Research*, 38 (2), 308–22.
- Parkman, John M. (1971), “Temporal aspects of digit and letter inequality judgments.,” *Journal of experimental psychology*, 91 (2), 191.
- Pinel, Philippe, Stanislas Dehaene, Denis Rivière, and Denis LeBihan (2001), “Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task,” *NeuroImage*, 14 (5), 1013–26.
- Read, Daniel, Shane Frederick, Burcu Orsel, and Juwaria Rahman (2005), “Four score and seven years from now: The date/delay effect in temporal discounting,” *Management Science*, 51 (9), 1326–35.
- Schley, Dan R. and Ellen Peters (2014), “Assessing ‘economic value’: Symbolic-number mappings predict risky and riskless valuations,” *Psychological Science*, 25 (3), 753–61.
- Sokolova, Tatiana (2023), “Days-of-the-Week Effect in Temporal Judgments,” *Journal of Consumer Research*, 50 (1), 167–89.

- Teeter, Adam (2020), “Ask Adam: Does Whiskey Age or Degrade in the Bottle After Opening?,” *VinePair*, (accessed March 11, 2023), [available at <https://vinepair.com/articles/does-whiskey-age-in-bottle-after-opening/>].
- Thomas, Manoj and Vicki Morwitz (2009a), “Heuristics in numerical cognition: Implications for pricing,” in *Handbook of pricing research in marketing*, Edward Elgar Publishing, 132–49.
- Thomas, Manoj and Vicki Morwitz (2009b), “The Ease-of-Computation Effect: The Interplay of Metacognitive Experiences and Naive Theories in Judgments of Price Differences,” *Journal of Marketing Research*, 46 (1), 81–91.
- Ülkümen, Gülden and Manoj Thomas (2013), “Personal relevance and mental simulation amplify the duration framing effect,” *Journal of Marketing Research*, 50 (2), 194–206.
- Walsh, Vincent (2003), “A theory of magnitude: common cortical metrics of time, space and quantity,” *Trends in Cognitive Sciences*, 7 (11), 483–88.
- Weber, Ernst Heinrich (1834), *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae ...*, C.F. Koehler.
- Zauberman, Gal, B. Kyu Kim, Selin A. Malkoc, and James R. Bettman (2009), “Discounting Time and Time Discounting: Subjective Time Perception and Intertemporal Preferences,” *Journal of Marketing Research*, 46 (4), 543–56.
- Zauberman, Gal, Jonathan Levav, Kristin Diehl, and Rajesh Bhargave (2010), “1995 feels so close yet so far: The effect of event markers on subjective feelings of elapsed time,” *Psychological Science*, 21 (1), 133–39.
- Zauberman, Gal, Rebecca K. Ratner, and B. Kyu Kim (2009), “Memories as assets: Strategic memory protection in choice over time,” *Journal of consumer research*, 35 (5), 715–28.

Zbrodoff, N. Jane and Gordon D. Logan (2005), “What Everyone Finds: The Problem-Size Effect,” in *The Handbook of Mathematical Cognition*, Psychology Press.

Zhang, Y. Charles and Norbert Schwarz (2012), “How and why 1 year differs from 365 days: A conversational logic analysis of inferences from the granularity of quantitative expressions,” *Journal of Consumer Research*, 39 (2), 248–59.

CHAPTER 1 APPENDIX

Fooled by Stars: Misestimations in Judgments of Numeric Ratings

Contents

1. Table W1: Popular E-Commerce Platforms-Rating Symbols on Product Search Page
2. Figure W1
3. Supplementary Study 1
4. Supplementary Study 2
5. Supplementary Study 3
6. Study 1 Stimuli, Data Exclusion, and modified Figure 3A
7. Study 2 Stimuli and Data Exclusion
8. Study 3 Stimuli, Data Exclusion, and Figure W5
9. Study 4 Stimuli and Data Exclusion
10. Study 5 Stimuli, Data Exclusion, and Figure W6

Please note that the images included in this appendix are not to scale and are smaller in size than the stimuli used in the experiments.

Table W1: Popular E-Commerce Platforms - Rating Symbols on Product Search Page

S. No.	Name	Description	Rating Symbol on Search Page
1	Target	Major US retailer known for a wide assortment of everyday goods. This retailer uses star rating symbols.	 Link
2	TripAdvisor	Internationally recognized U.S. travel review and booking platform. This retailer uses circle rating symbols.	 Link
3	The Home Depot	Premier U.S. home improvement retailer. This retailer uses star rating symbols.	 Link
4	Sephora	Prominent U.S. beauty and personal care retailer. This retailer uses star rating symbols.	 Link
5	Macy's	Renowned U.S. department store brand. This retailer uses star rating symbols.	 Link
6	Best Buy	Top electronics retailer in North America. This retailer uses star rating symbols.	 Link
7	Lowe's	Leading competitor in U.S. home improvement. This retailer uses star rating symbols.	 Link
8	Costco	Popular bulk retailer with online shopping. This retailer uses star rating symbols.	 Link
9	Newegg	Online electronics and computer retailer. This retailer uses egg shaped rating symbols.	 Link
10	Uber Eats	Global food delivery platform. This platform uses Arabic numeral rating symbols.	4.5 Link
11	Airbnb	A marketplace for lodging and tourism experiences. This platform uses Arabic numeral rating	★ 4.89 Link

		symbols.	
12	Rotten Tomatoes	A review aggregation website for movies and TV shows. This platform uses Arabic numeral rating symbols.	 79%  91% Link
13	Glassdoor	A review platform for employees. This platform uses Arabic numeral rating symbols.	4.3 ★ Link
14	Doordash	US based food delivery service. This platform uses Arabic numeral rating symbols.	4.6 ★ Link
15	Expedia	Global online travel agent. This platform uses Arabic numeral rating symbols.	7.4 Link
16	Postmates	Food delivery platform. This platform uses Arabic numeral rating symbols.	4.5 Link
17	Yelp	Top global online consumer review platform. This platform uses both star and Arabic numeral rating symbols.	 4.3 Link
18	Google Maps/Places	Global local business directory. This platform uses both star and Arabic numeral rating symbols.	4.3 ★★★★★ Link
19	Goodreads	Online book recommendation community. This platform uses both star and Arabic numeral rating symbols.	 4.47 Link
20	Trustpilot	Consumer reviews for businesses worldwide. This platform uses both star and Arabic numeral rating symbols.	 TrustScore 4.4 Link
21	Apple's App Store	Apple's digital app and software marketplace. This platform uses both star and Arabic numeral rating symbols.	4.6 ★★★★★ Link

Table W1— This table shows how popular e-commerce platforms differ in their choice of rating symbols on the product search page.

Figure W1: Distribution of Average Product Ratings for Snacks on Amazon USA

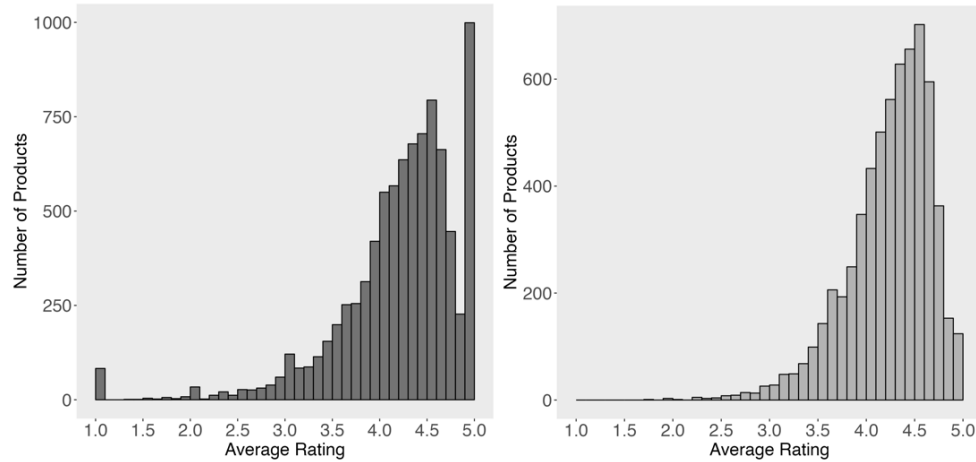
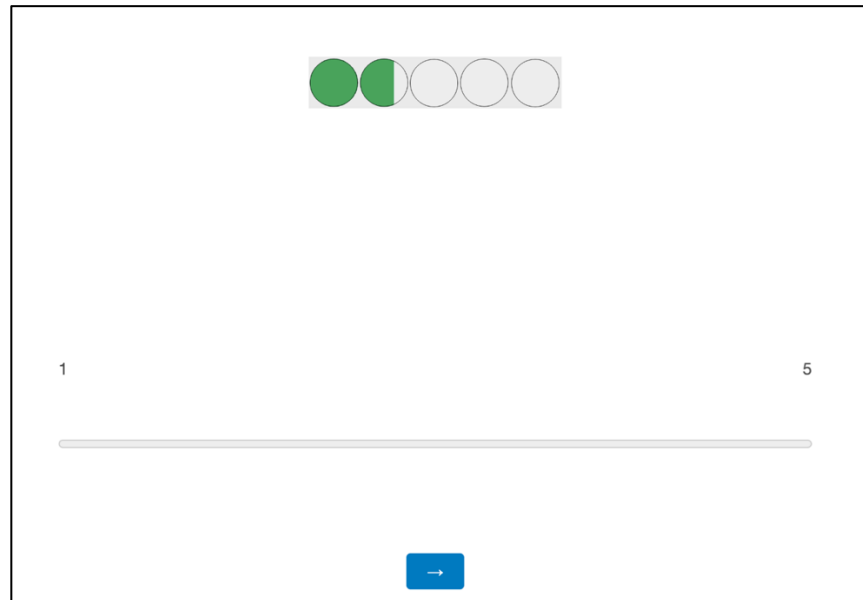


Figure W1— The left panel illustrates the distribution of product ratings in the snack category on Amazon USA in 2019. The right panel refines this distribution to include only products with 10 or more ratings.

SUPPLEMENTARY STUDY 1: CIRCLES VERSUS ARABIC NUMERALS

The instructions and paradigm of this study was the same as Study 1. Following is one of the examples of the 17 questions administered in the estimation task in the circle condition.



Supplementary Study 1 was designed to determine whether the overestimation of graphically represented magnitudes is unique to star-shaped symbols. It is conceivable that the

overestimation seen in Study 1 could be attributed to the positive associations elicited by stars. If that were the case, we would expect this effect not to extend to other, less favored graphical symbols such as circles. However, according to our theory, the overestimation of graphical ratings results from visual completion, and thus even an incomplete circle would be overestimated. To test this hypothesis, we employed a similar paradigm to that used in Study 1 to compare participants' magnitude judgments of numeric ratings represented using circles and Arabic numerals. This study was pre-registered at OSF (<http://bit.ly/3YvQHu2>).

Participants and Procedure

We recruited 452 participants using a US standard sample from Prolific in exchange for monetary compensation. We had pre-registered to recruit 450 participants, but we found two additional participants in the Qualtrics dataset. This could have happened because few subjects might have failed to enter the correct completion code in the end but had nonetheless completed the survey. Hence, these participants weren't counted by CloudResearch but their data were registered in Qualtrics. All our results hold if we only include the first 450 participants (as per the chronological order). Per our pre-registered exclusion criteria, we removed any participant who met any of the following conditions: duplicate IP addresses ($n = 2$) or response time to complete the survey was three standard deviations above or below the mean response time ($n = 7$). We analyzed the data from 443 participants (43.1% female, 55.3% male, 1.5% nonbinary, $M_{\text{age}} = 38$ years).

In this study, we employed the same procedure as in Study 1, but with two modifications. First, we substituted stars with circles, and second, we did not use the dual representation condition. Participants were randomly assigned to either the circle or Arabic numeral rating

symbol condition. They underwent a calibration phase and then completed the mental number line task, followed by standard demographic questions on age, gender, income, and education level.

Analyses and Results

The findings of this study largely align with those of Study 1 (See Figures W2A and W2B). To analyze the data, we calculated the deviation between the perceived magnitude of the ratings and their actual magnitude. We followed a similar approach to Study 1 and averaged the deviations for whole ratings, fractions with quarters, fractions with halves, and fractions with three-fourths. We then used a 2 x 4 ANOVA with the rating symbol (circle vs. Arabic numeral) as a between-subjects factor and number type (whole, quarter, half, and three-fourths) as a within-subject factor¹⁴. Our analysis revealed a significant main effect of rating symbol ($F(1, 441) = 34.72, p < .0001, \eta^2_p = .07$) and a significant two-way interaction ($F(3, 1323) = 24.27, p < .0001, \eta^2_p = .05$).

¹⁴ We pre-registered a 2 x 2 mixed ANOVA analysis with rating symbol (star, Arabic numeral) as the between-subjects factor and number type (whole, fractional) as the within-subjects factor. All our results hold if we analyze the data using our pre-registered ANOVA model.

Figure W2A: Perceived Magnitude of Ratings

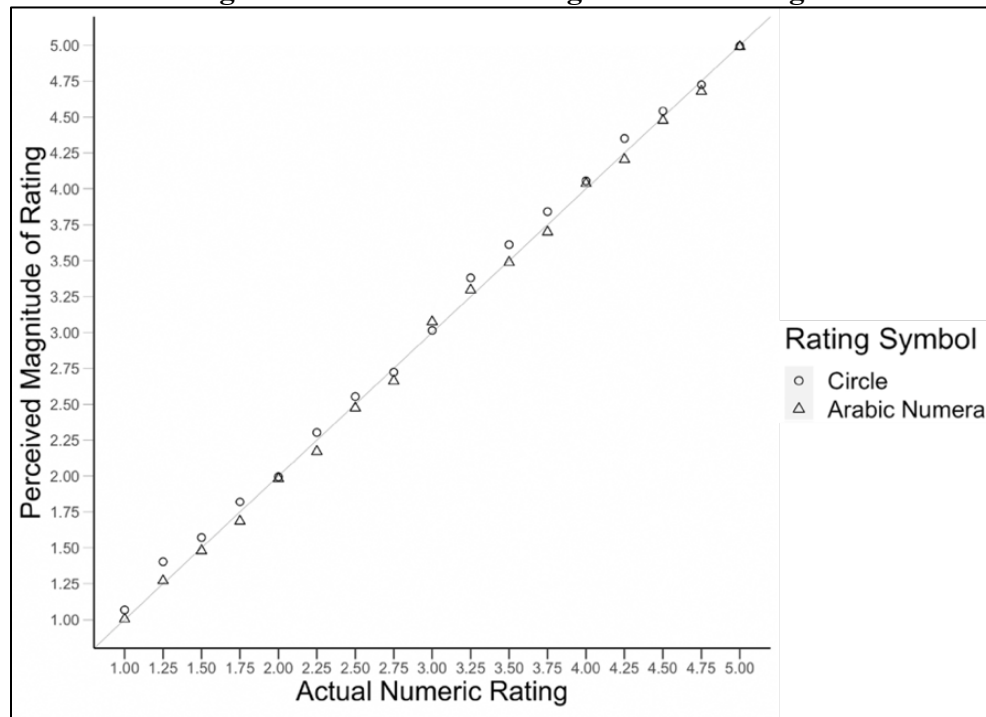
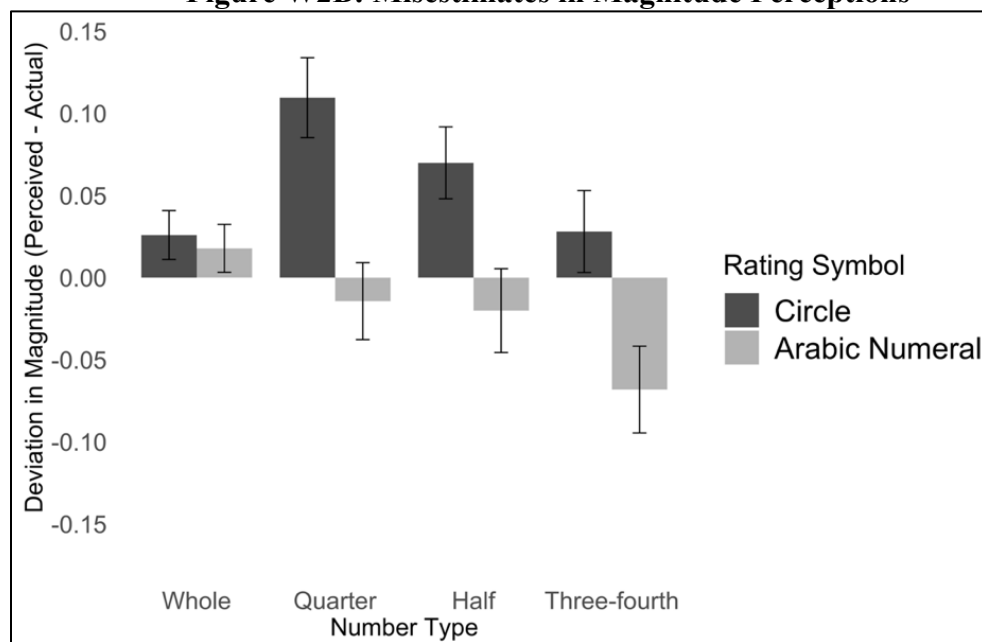


Figure W2B: Misestimates in Magnitude Perceptions



These figures show the results from Supplementary Study 1. The top figure shows how the average perceived magnitudes of the 17 stimuli change with different rating symbols. The bottom figure shows the misestimate in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) by type of number across rating symbols. The bars in bottom figure represent 95% confidence intervals. Both figures depict raw means.

First, participants overestimated fractional circle ratings ($M_{\text{Circle}} = +.07 [+0.05, +.09]$), but underestimated fractional Arabic numeral ratings on average¹⁵ ($M_{\text{ArabicNumeral}} = -.03 [-.05, -.015]$; $F(1, 483) = 56.0, p < .0001, \eta^2_p = .10$); supporting H1a and H1b. Second, supporting H2, we found that the overestimation in the perception of circle ratings was more pronounced for incomplete circles ($M_{\text{FractionalCircles}} = +.07 [+0.05, +.09]$) than for complete circles ($M_{\text{WholeCircles}} = +.03 [+0.00, +.05]$; $F(1,1323) = 27.60, p < .0001, \eta^2_p = .02$). Similarly, the underestimation in the perception of Arabic numeral ratings was greater for fractional numeric ratings ($M_{\text{FractionalArabicNumerals}} = -.03 [-.05, -.01]$) than for whole number ratings ($M_{\text{WholeArabicNumerals}} = +.02 [-.00, +.04]$; $F(1,1323) = 39.36, p < .0001, \eta^2_p = .03$).

Third, the overestimation in the perception of fractional circle ratings increased when the anchor was farther from the actual number. The overestimation was higher for quarter ratings ending in .25 ($M_{.25\text{Circle}} = +.11 [+0.09, +.13]$) than for three-fourths ratings ending in .75 ($M_{.75\text{Circle}} = +.03 [+0.01, +.05]$; $F(1,1323) = 65.25, p < .0001, \eta^2_p = .05$). This supports H4a. The pattern reversed for Arabic numeral ratings; the underestimation was more pronounced for three-fourths ratings ending in .75 ($M_{.75\text{ArabicNumerals}} = -.07 [-.09, -.05]$) than for quarter ratings ending in .25 ($M_{.25\text{ArabicNumerals}} = -.01 [-.04, +.01]$; $F(1,1323) = 28.13, p < .0001, \eta^2_p = .02$); this supports H4b.

Discussion

Supplementary Study 1 reinforces the findings observed in Study 1, as illustrated in the figures W2A and W2B; supporting H1a, H1b, H2, H4a, and H4b. To further explore the generalizability of these results, we conducted Supplementary Study 2, utilizing squares as

¹⁵ The underestimation for quarter and half Arabic numerals was not statistically significant. H4b predicts how the magnitude of underestimation varies across different fractional numbers, which accounts for this observation.

graphical symbols. This study replicated the patterns observed with stars and circles, highlighting a consistent overestimation in magnitude judgments when numerical quantities are represented by various graphical symbols, including stars, circles, and squares.

SUPPLEMENTARY STUDY 2: SQUARES

In Supplementary Study 2, our objective was to explore the presence of overestimation in graphical ratings when square symbols are utilized to represent magnitudes. Existing research, such as the work by Jia, Wan, and Zheng (2023), posits that the principle of visual completion might not extend to rectangular shapes, including squares. This is based on the premise that a square, even when only partially filled, might still be perceived as complete geometric entity, such as rectangle. For visual reference, see the stimuli provided below. To examine this hypothesis, we conducted an empirical study focusing on square symbols.

Stimuli used to show magnitude of 3.25 using square symbols



We used the same experimental paradigm used in Study 1 and Supplementary Study 1 for Supplementary Study 2. However, in this study, our focus was specifically on the accuracy of magnitude judgments using square rating symbols. Consequently, we employed only one type of rating symbol. Participants were presented with 17 square ratings ranging from 1 to 5, increasing in increments of 0.25, in a randomized sequence.

We recruited 38 participants (34% non-male, $M_{\text{age}} = 37$ years) from a US standard sample via Prolific, offering monetary compensation for their participation. The analytical approach mirrored that of Study 1 and Supplementary Study 1. We calculated the deviation between the perceived and actual magnitudes of the ratings. Deviations were averaged across different types

of ratings: whole numbers, fractions with quarters, halves, and three-fourths. A repeated measures ANOVA model was used, with number type (whole, quarter, half, and three-fourths) as the within-subject factor.

Our analysis yielded a significant main effect for number type ($F(3, 111) = 8.09, p = .0001, \eta^2_p = .18$). Consistent with previous findings, fractional ratings were overestimated while the misestimation in whole ratings was not significantly different from zero ($M_{\text{Fractional}} = +.11 [+.07, +.16], M_{\text{Whole}} = +.035 [-.02, +.09]; F(1, 111) = 15.10, p = .0002, \eta^2_p = .12$). Within the fractional ratings, the overestimation was most pronounced for quarter ratings and least for three-fourth ratings ($M_{.25} = +.15 [+.09, +.20], M_{.75} = +.07 [+.02, +.13]; F(1, 111) = 8.94, p = .003, \eta^2_p = .07$). The deviation for half ratings was found to be $+.12 [+.07, +.17]$.

In conclusion, Supplementary Study 2 validated the robustness of our earlier findings, demonstrating that the overestimation in magnitude judgments transcends different graphical symbols — be it stars, circles, or squares.

SUPPLEMENTARY STUDY 3: FILL METHOD USED FOR FRACTIONAL STARS

There are two different methods to fill fractional stars—area-proportion fill and linear-proportion-fill-on-width. Does the fill method influence the magnitude perception of fractional stars?

To visually express a fraction, like 0.2, using a star, designers must divide the star vertically into two segments: the left filled in, indicating the fraction, and the right remaining empty. There are two methods to determine where to make this split. In the ‘area-proportion’ method, the filled left portion represents 20% of the star’s total area, creating an area-proportional representation of the fraction. Alternatively, the ‘linear-proportion-fill-on-width’ involves cutting the star such that 20% of the width length starting from the left-edge is filled.

The non-rectangular shape of the star results in different visual representations of fractional values when employing area-proportion and linear-proportion-on-width fill methods. For fractions less than 0.5, the area-proportion method results in more filling (i.e., more ink), whereas for fractions above 0.5, the linear-proportion-on-width method leads to more filling. Such differences are critical in marketing, where non-rectangular symbols like stars and circles are common in depicting ratings. The selected method of fill can thus sway consumers' perceptions of a product's magnitude rating. This study delves into how these variations in star fill influences consumer judgment. The Figure W3 offers a visual comparison of the filling discrepancies across the two methods for different fractional values.

Figure W3: Comparative Visualization of Fractional Ratings Using Area-Proportion and Linear-Proportion-on-Width Fill Methods

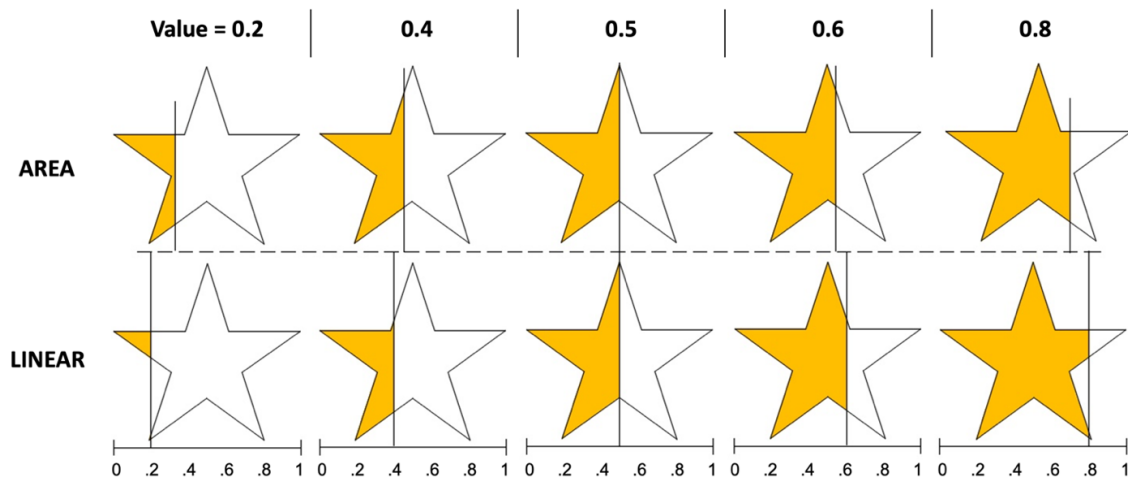


Figure W3—This figure displays the area-proportion and the linear-proportion-on-width fill methods applied to star symbols to represent fractional values of 0.2, 0.4, 0.5, 0.6, and 0.8. The left-hand side of each star is shaded to correspond with the fractional value. The top panel shows stars with area-proportion fill while the bottom panel shows stars with linear-proportion-on-width fill. The amount of fill (i.e., the ink) is greater in the area-proportion stars compared to linear-proportion-on-width stars for fractions below 0.5, while this pattern reverses for fractions above 0.5.

The question—whether area-proportion method and linear-proportion-on-width method have different effects on magnitude perception—is particularly important given the recent findings reported by Jia, Wan, and Zheng (2023). The authors compared stars with bars and found that star ratings are overestimated compared to bar ratings for fractional values above 0.5, while they are underestimated for values below 0.5. We believe that this pattern of results might be contingent on the fill method used for the graphic ratings. In all our studies, we used the area-proportion method to create our stimuli, following the guidelines in the field of data visualization (Bergstrom and West 2016; Tufte, 2001)¹⁶. We found that stars created using the area-proportion

¹⁶ In computer graphics, to prevent deceptive visual data representations, the ‘principle of proportional ink’ mandates that the *area* used in graphs to represent data should correspond precisely to the data values (Bergstrom and West 2016). This concept is derived from Edward Tufte’s classic book ‘The Visual Display of Quantitative Information’, where he argues that “The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented” (Tufte, 2001). Furthermore, in stars, while area filling ensures a consistent visual increment of ink for proportional value increases (e.g., the ink increment between 0.1 to 0.2 is the same as between 0.4 and 0.5), linear filling results in disproportionate visual increments of ink for equivalent numerical increases, leading to potential deceptive interpretations (e.g., the ink increment between 0.1 to 0.2 is considerably less than between 0.4 and 0.5).

method are always overestimated. However, we speculated that using the linear-proportion-on-width method might yield a different pattern of results, aligned with those reported by Jia et al. The present study was designed to answer this question (pre-registered at OSF; <https://bit.ly/49Vd3L8>).

Participants and Procedure

We recruited 349 participants using a US standard sample from Prolific in exchange for monetary compensation. Per our pre-registered exclusion criteria, we removed participants whose completion times were three standard deviations from the mean ($n = 6$). We analyzed the data from 343 participants (49.6% female, 48.7% male, 1.7% nonbinary, $M_{\text{age}} = 41$ years). The study followed a number line estimation paradigm similar to that of Study 1 and Supplementary Study 1, but with two key modifications. Firstly, participants were randomly assigned to one of two conditions, each featuring a different type of star symbol. One condition utilized stars filled using the area-proportion method, while the other used the linear-proportion-on-width method. Secondly, we made an adjustment to the rating magnitudes: instead of using quarter and three-quarter ratings, we included ratings ending in .2 and .8 to employ the same rating values as used in Jia et al¹⁷. This modification resulted in a set of 17 distinct rating magnitudes ranging from 1, 1.2, 1.5, 1.8, 2, 2.2, 2.5, 2.8, through to 5.

Analyses and Results

As in the previous studies, we calculated the deviation between perceived and actual

¹⁷ The star rating stimuli used in the linear condition were the exact ones utilized by Jia, Wan, and Zheng (2023). However, they did not include linear filled quarter and three-quarter star stimuli, and hence we substituted these with 0.2 and 0.8 rating magnitudes for the purpose of our study.

magnitude ratings. After averaging these deviations for each number type, we conducted a 2 x 4 mixed ANOVA, with the type of star fill method (area vs. linear) as the between-subjects factor and number type (whole numbers, and those ending in .2, .5, and .8) as the within-subjects factor. The main effect of number type was significant ($F(3, 1023) = 25.12, p < .0001, \eta^2_p = .07$), and so was the interaction effect between star fill method and number type ($F(3, 1023) = 59.16, p < .0001, \eta^2_p = .15$). The main effect for the star fill method was not significant ($F(1, 341) = 2.01, p = .16, \eta^2_p = .006$). See Figure W4 for a visual depiction of these results.

Figure W4: Impact of Fill Method on Star Rating Magnitudes

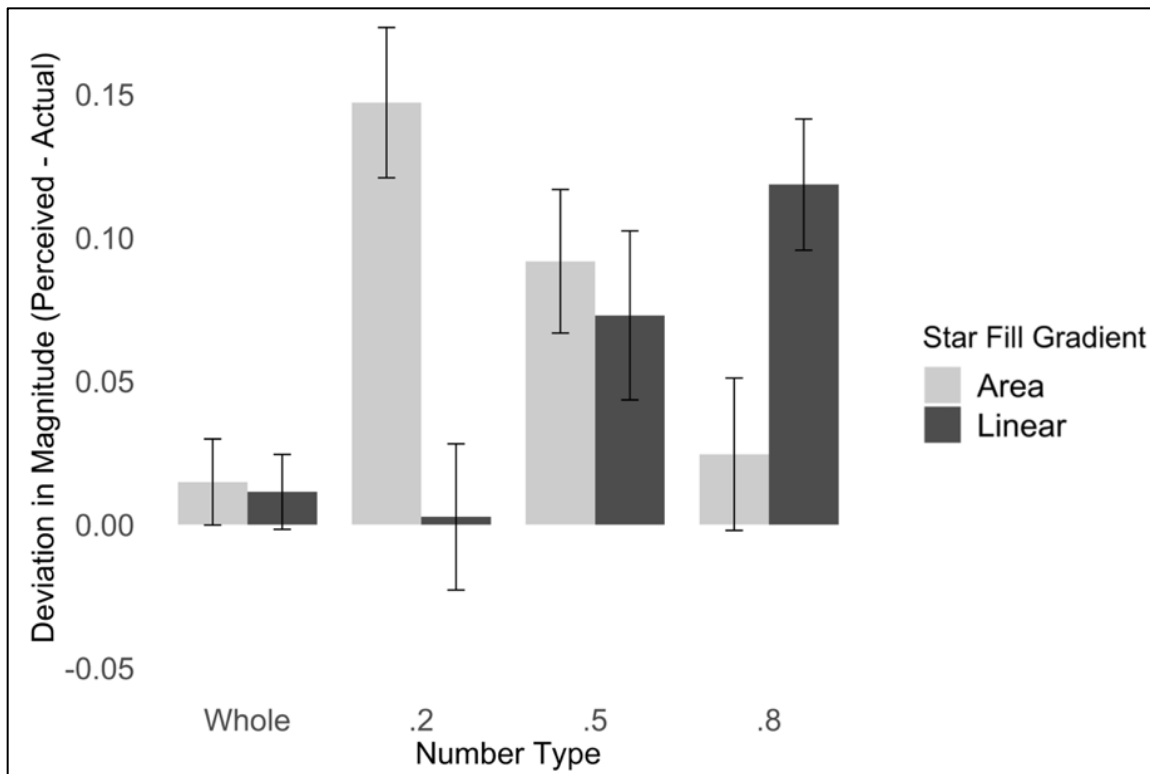


Figure W4—This figure shows the results of Supplementary Study 3, illustrating the misestimations in subjective magnitude perception for star ratings—calculated as the difference between perceived and actual magnitudes. The displayed misestimations pertain to stars filled via either the area-proportion method or the linear-proportion-on-width method. Bars indicate 95% confidence intervals, and the plot presents raw mean values.

For numbers ending in .2, the overestimation was more when the stars were created using

the area fill than for stars with linear fill method ($M_{\text{Area_Stars}} = +.15$ [$+.12, +.17$], $M_{\text{Linear_Stars}} = +.00$ [$-.02, +.03$]; $F(1, 873) = 73.02, p < .0001, \eta^2_p = .08$). In contrast, for numbers ending in .8, the overestimation was more in the linear fill method ($M_{\text{Area_Stars}} = +.02$ [$+.00, +.05$], $M_{\text{Linear_Stars}} = +.12$ [$+.10, +.14$]; $F(1, 873) = 30.92, p < .0001, \eta^2_p = .03$). As expected, the deviation for whole numbers ($M_{\text{Area_Stars}} = +.01$ [$-.01, +.03$], $M_{\text{Linear_Stars}} = +.01$ [$-.01, +.03$]; $F(1, 873) = .04, p = .84, \eta^2_p = .00$) and for numbers ending in .5 ($M_{\text{Area_Stars}} = +.09$ [$+.07, +.12$], $M_{\text{Linear_Stars}} = +.07$ [$+.05, +.10$]; $F(1, 873) = 1.24, p = .27, \eta^2_p = .00$) was not significantly different between the two star fill methods, which aligns with the expectation as these conditions do not differ visually in their representation.

Discussion

The findings from Supplementary Study 3 provide compelling evidence that the fill method employed significantly shapes consumer perceptions of star ratings, particularly for non-half fractional values. Due to the two fill methods resulting in differing amounts of fill, area filled stars were overestimated for fractions below 0.5, while linear filled stars were overestimated for fractions above 0.5.

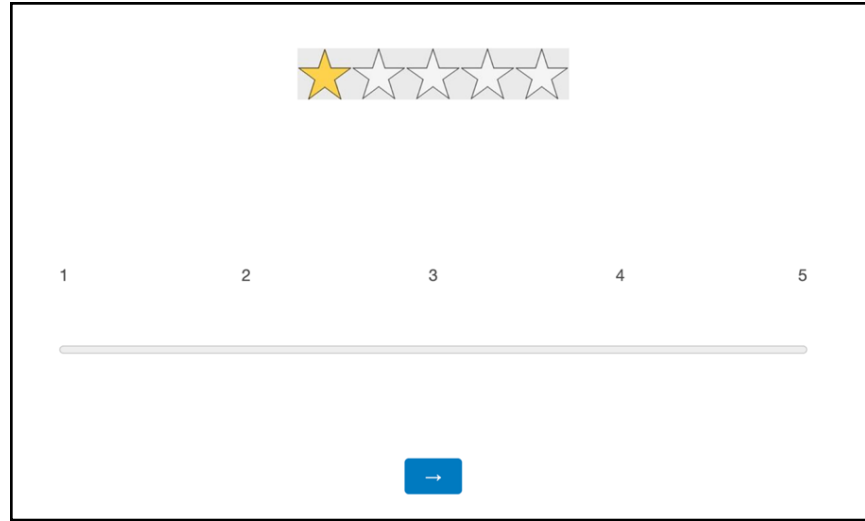
These patterns mirror those observed by Jia, Wan, and Zheng (2023) in their study comparing linear filled stars with bars. Because bars are rectangular, the linear and area fill methods both yield an area-proportional representation of the fraction. Thus, their study effectively juxtaposed two graphical symbols distinguished by their method of fill—one with linear and the other with area filling. Notably, for fractions less than 0.5, linear filled stars had less fill than area filled bars, whereas for fractions greater than 0.5, linear filled stars displayed more fill. This parallel suggests that the principal insights from Jia, Wan, and Zheng's (2023)

study are more likely a result of the linear fill method employed in the stars, rather than intrinsic differences between the star and bar shapes.

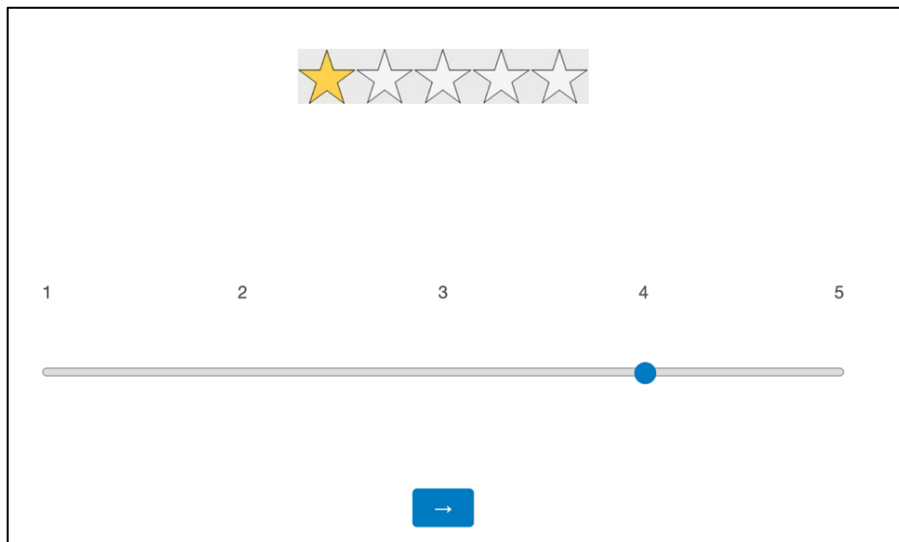
STUDY 1 STIMULI, DATA EXCLUSION, AND MODIFIED FIGURE 3A

1. Calibration Phase– Star Rating Condition

Participants were administered three questions in the calibration phase. Following is the first question. It was followed by two more questions for values of 3 and 5.



Participants could respond by clicking on the scale. The starting position of the cursor was hidden. The cursor would appear on the scale at the position of the first click. Once the cursor appeared on the scale, participants could move the cursor by dragging it along the scale.



An error message appeared if participant submitted an incorrect response, and the participant

was asked to give their response again until they responded correctly.

Sorry, this is not the correct response. Please try again



1 2 3 4 5

→

2. Estimation Task

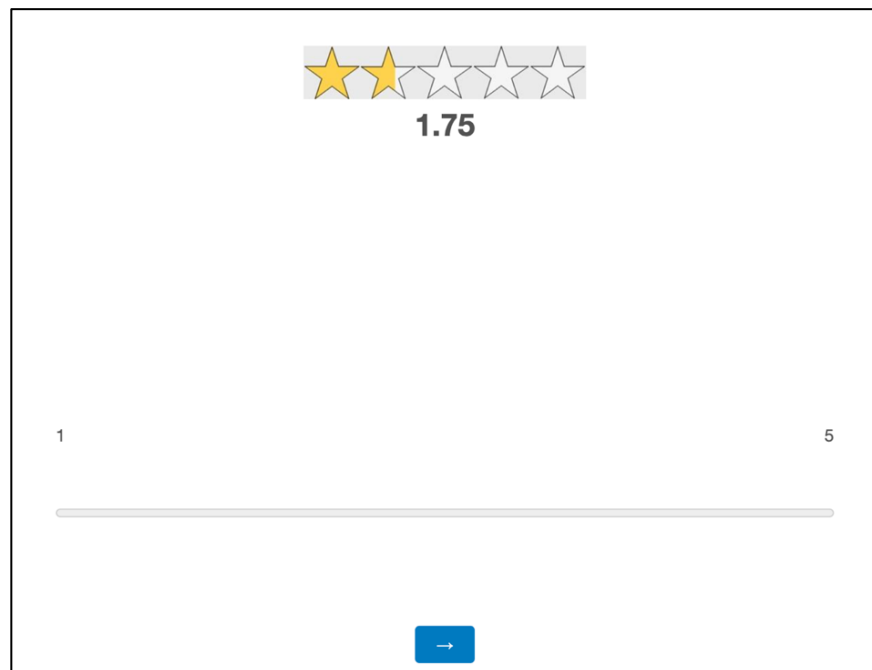
Participants were administered 17 questions in the estimation task phase. The order of 17 questions was randomized. Each question was displayed on a separate screen.

Star Rating Condition (rating = 1.75)



The interface for the Star Rating Condition shows a horizontal row of five stars at the top. The first two stars are filled with yellow, and the remaining three are white with black outlines. Below the stars, a horizontal slider bar is positioned, with the number '1' at the left end and '5' at the right end. A blue button with a white right-pointing arrow is located at the bottom center of the interface.

Star + Arabic Numeral Condition (rating = 1.75)

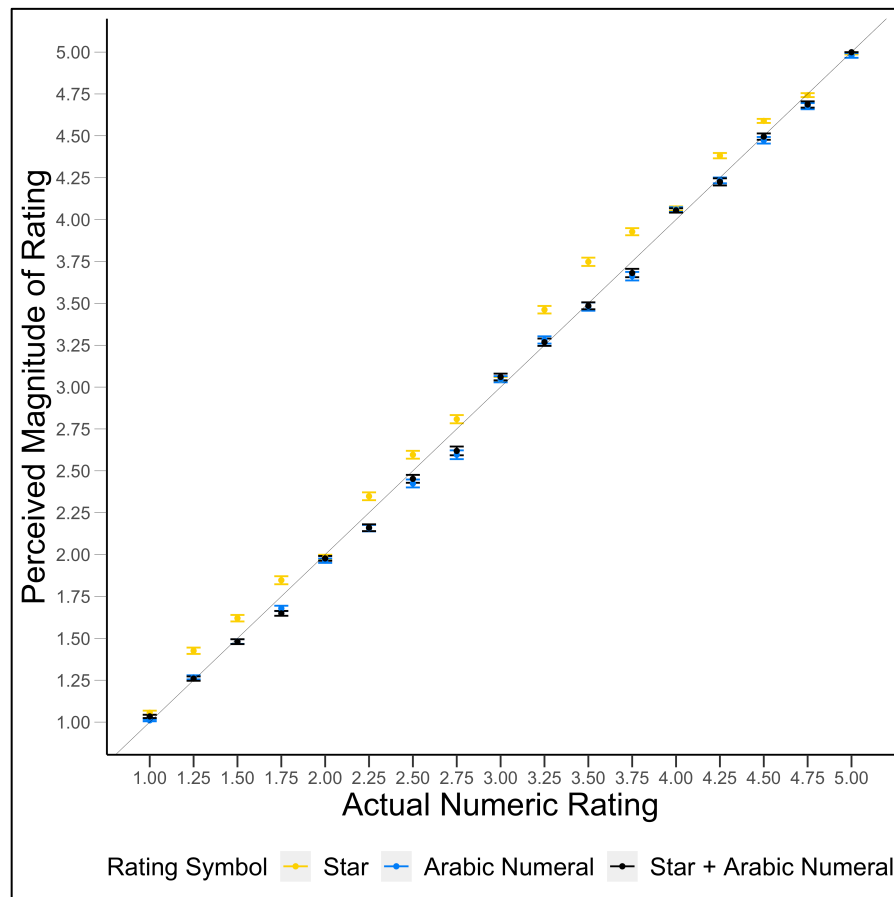


The interface for the Star + Arabic Numeral Condition is similar to the Star Rating Condition, but it includes the Arabic numeral '1.75' centered below the row of five stars. The first two stars are filled with yellow, and the remaining three are white with black outlines. Below the stars, a horizontal slider bar is positioned, with the number '1' at the left end and '5' at the right end. A blue button with a white right-pointing arrow is located at the bottom center of the interface.

Arabic Numeral Condition (rating = 1.75)



Modified Figure 3A (with 95% Confidence Intervals)



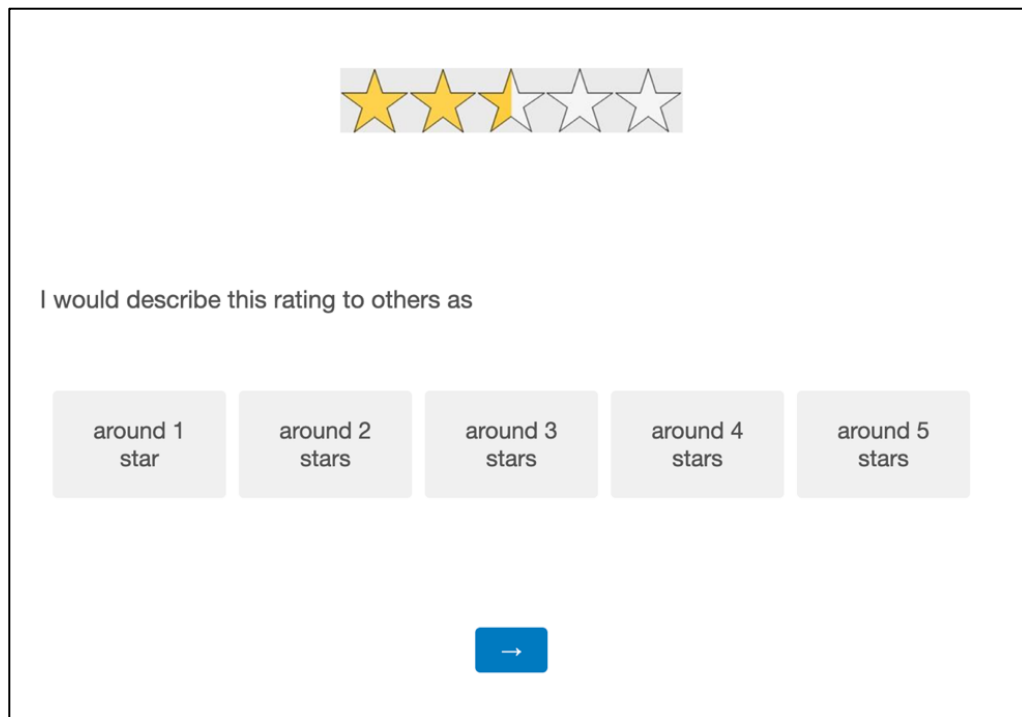
Data Exclusion

We recruited 624 participants using a US standard sample from Prolific in exchange for monetary compensation. We had pre-registered to recruit 625 participants, but we found one participant less in the Qualtrics dataset due to a technical error. Per our pre-registered exclusion criteria, we removed eight participants because their response time to complete the survey was three standard deviations above or below the mean response time. We analyzed the data from 616 participants (45% non-male, $M_{\text{age}} = 41$ years).

STUDY 2 STIMULI AND DATA EXCLUSION

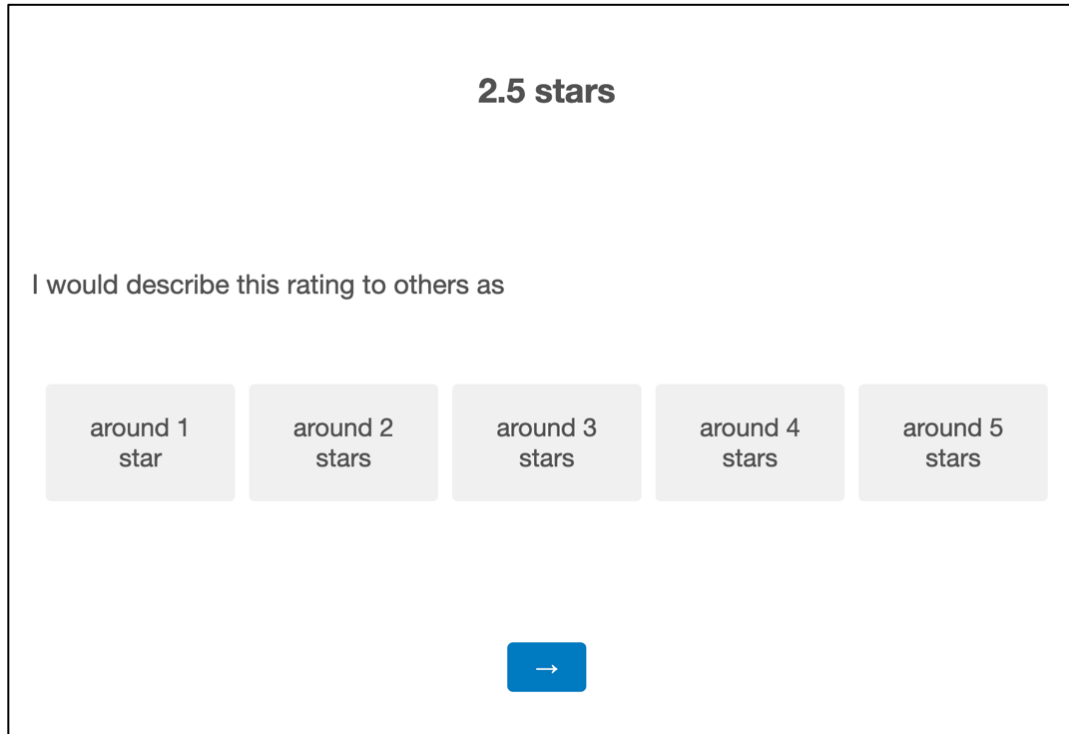
Participants answered four questions in each condition. The order of the four questions was randomized. Each question was displayed on a separate screen.

Star Rating Condition (rating = 2.5)



The interface for the Star Rating Condition is displayed within a rectangular frame. At the top, there is a horizontal row of five stars. The first three stars are filled with a yellow color, and the last two stars are white with a gray outline. Below the stars, the text "I would describe this rating to others as" is centered. Underneath this text, there are five rectangular buttons arranged horizontally. Each button contains the text "around [number] stars", where the number is 1, 2, 3, 4, or 5. At the bottom center of the frame, there is a blue rectangular button with a white right-pointing arrow.

Arabic Numeral Condition (rating = 2.5)



2.5 stars

I would describe this rating to others as

around 1 star around 2 stars around 3 stars around 4 stars around 5 stars

→

Data Exclusion

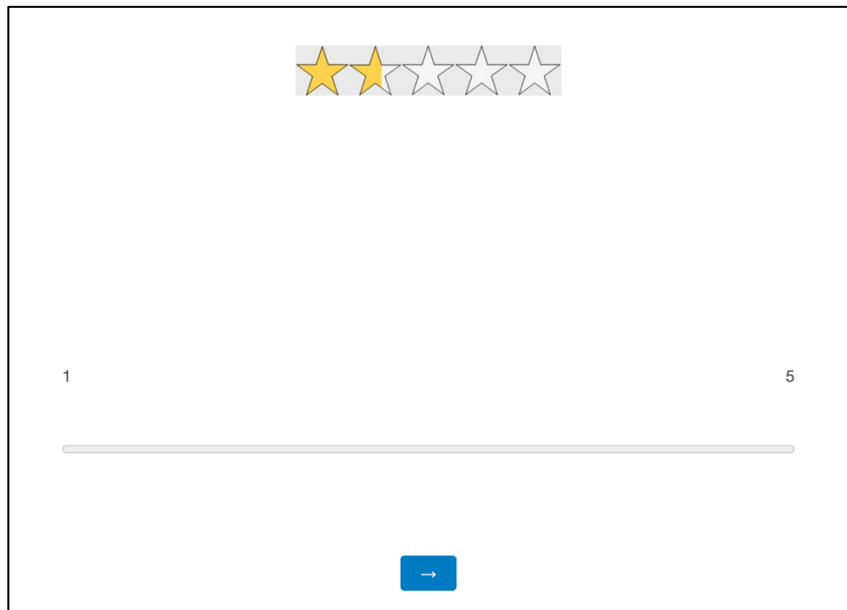
We recruited 327 participants using the CloudResearch approved sample from the CloudResearch panel of US participants in exchange for monetary compensation. We had pre-registered to recruit 325 participants, but we found two additional participants in the Qualtrics dataset. All our results hold if we only include the first 325 participants (as per the chronological order). Per our pre-registered exclusion criteria, we removed six participants because their response time to complete the survey was three standard deviations above or below the mean response time. This left us with 1284 choice responses from 321 participants (41% non-male, $M_{\text{age}} = 40$ years). We removed choice responses in which neither the rounded-up nor rounded-down rating value was chosen ($n = 7$), and finally analyzed the data of 1277 choice responses.

STUDY 3 STIMULI, DATA EXCLUSION, AND FIGURE W5

The instructions and paradigm of Study 3 was the same as Study 1.

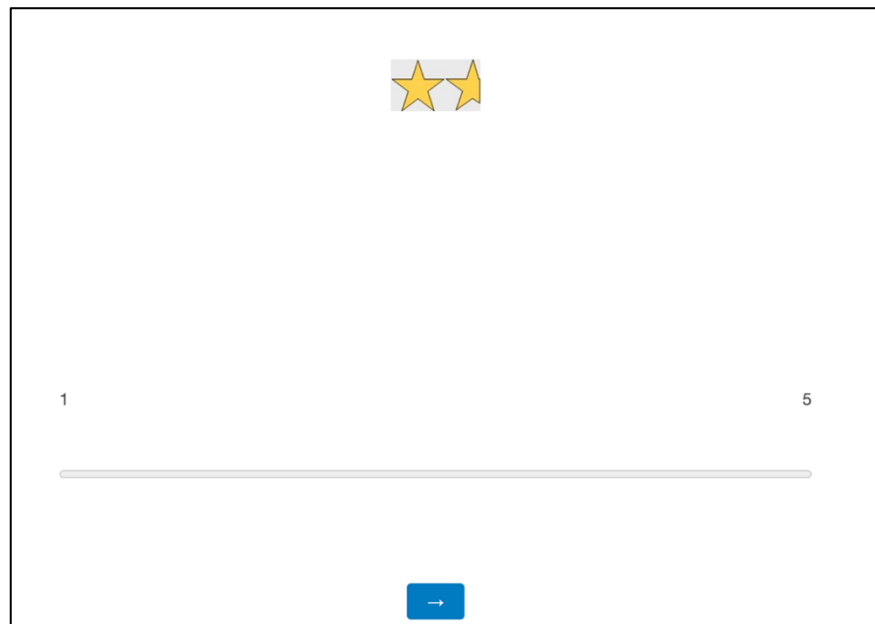
Following are examples of questions administered in the estimation task in the visually incomplete and complete star conditions.

Visually incomplete star condition (rating = 1.75)



The interface for the visually incomplete star condition shows a row of five stars at the top. The first two stars are yellow, and the remaining three are white with black outlines. Below the stars is a horizontal slider bar with the number '1' at the left end and '5' at the right end. A blue button with a right-pointing arrow is centered below the slider bar.

Visually complete star condition (rating = 1.75)



The interface for the visually complete star condition shows a row of two yellow stars at the top. Below the stars is a horizontal slider bar with the number '1' at the left end and '5' at the right end. A blue button with a right-pointing arrow is centered below the slider bar.

Data Exclusion

We recruited 551 participants using a US standard sample from Prolific in exchange for monetary compensation. We had pre-registered to recruit 550 participants, but we found one additional participant in the Qualtrics dataset. All our results hold if we only include the first 550 participants (as per the chronological order). Per our pre-registered exclusion criteria, we removed eight participants because their response time to complete the survey was three standard deviations above or below the mean response time. We analyzed the data from 543 participants (55% non-male, $M_{\text{age}} = 42$ years).

Figure W5: Visual Completeness Mitigates Overestimation in Fractional Star Ratings

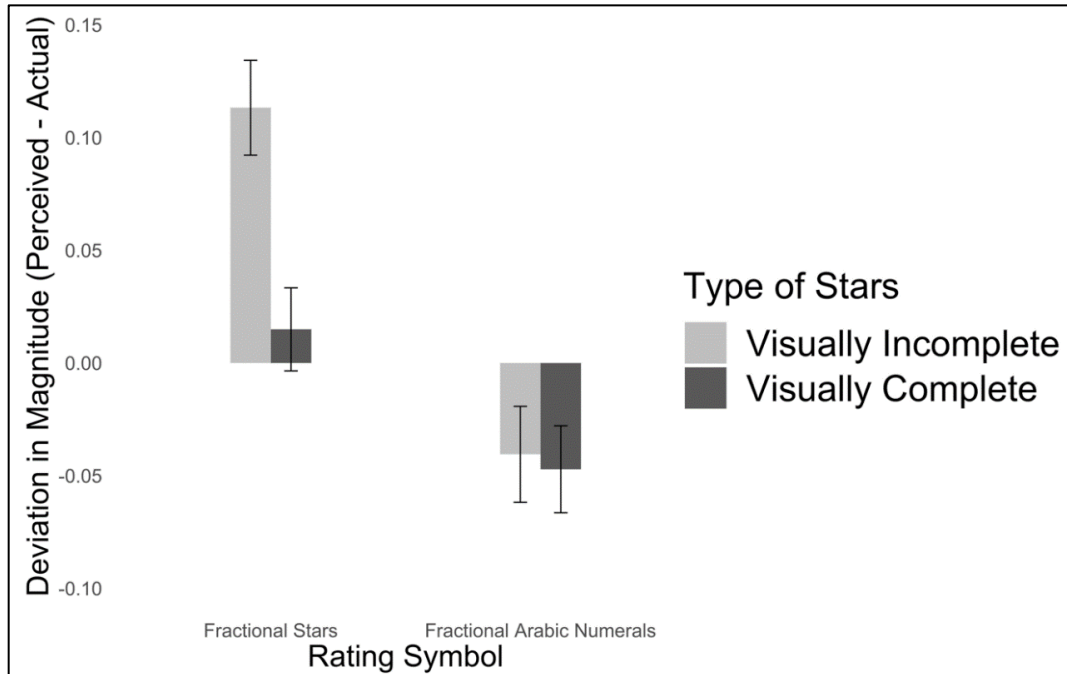


Figure W5—This figure shows the results of Study 3. The plot shows the misestimation in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) of fractional star ratings and fractional Arabic numeral ratings by type of stars used in Study 3. The bars represent 95% confidence intervals. The plot depicts raw means.

STUDY 4 STIMULI AND DATA EXCLUSION

Participants were administered four questions. The order of four questions was randomized. For each question, participants saw a perfume on first screen and then recorded their responses on the second screen. The second screen was the same in each of the four questions. The perfume brand and name were different in each question.

Second screen.

Please recall the name of the perfume as best as you can.

What is the rating of this perfume?

0 1 2 3 4 5 6

Rating

A horizontal slider bar with a blue dot positioned at the value 3.

→

Participants could move the cursor by dragging it. They could see the numeric value of the

position of the cursor as they dragged the cursor.

Please recall the name of the perfume as best as you can.

What is the rating of this perfume?

0 1 2 3 4 5 6

Rating

3.53

→

Star Rating Condition (rating =1.5), first screen



Arabic Numeral Condition (rating =1.5), first screen



Data Exclusion

We recruited 379 participants using the CloudResearch approved sample from the CloudResearch panel of US participants in exchange for monetary compensation. We had pre-registered to recruit 375 participants, but we found four additional participants in the Qualtrics dataset. All our results hold if we only include the first 375 participants (as per the chronological order). Per our pre-registered exclusion criteria, we removed any participant who met any of the following conditions¹⁸: duplicate IP addresses ($n = 4$), response time to complete the survey was three standard deviations above or below the mean response time ($n = 6$), or failed to verify that the product ratings were out of five stars ($n = 14$). This left us with 1420 responses from 355 participants for the analysis (53% non-male, $M_{\text{age}} = 39.6$ years). We further excluded 16 responses because they contained nonsensical or gibberish responses for the perfume name recalls, as specified in our pre-registered exclusion criteria. This was necessitated because we used open-ended text boxes to collect participants' responses in this study, unlike the previous studies where they responded using scales. We used 1404 recalled rating responses for the analysis.

¹⁸ We pre-registered to exclude recall responses that deviated by more than 1.5 points from the correct answer. Applying this criterion leads to exclusion of 18 participants, but it does not change our results. The results reported here do not exclude these participants.

STUDY 5 STIMULI, DATA EXCLUSION, AND FIGURE W6

Participants answered four WTP questions. The order of the four questions was randomized. Each question was displayed on a separate screen.

Star Rating Condition (rating = 1.5)

PILOT The Better Ballpoint Pen
Rating: ★★★★★

How much would you be willing to pay (\$1 to \$10)?

Arabic Numeral Condition (rating = 1.5)

PILOT The Better Ballpoint Pen
Rating: 1.5 ★

How much would you be willing to pay (\$1 to \$10)?

Next, participants answered four quality judgment questions. The order of the four questions was randomized. Each question was displayed on a separate screen.

Star Rating Condition (rating = 2.5)

Paper Mate Profile Ballpoint Pen
Rating: ★★★★★

How would you rate the quality of the pen?

Very Low

Low

High

Very High

Arabic Numeral Condition (rating = 2.5)

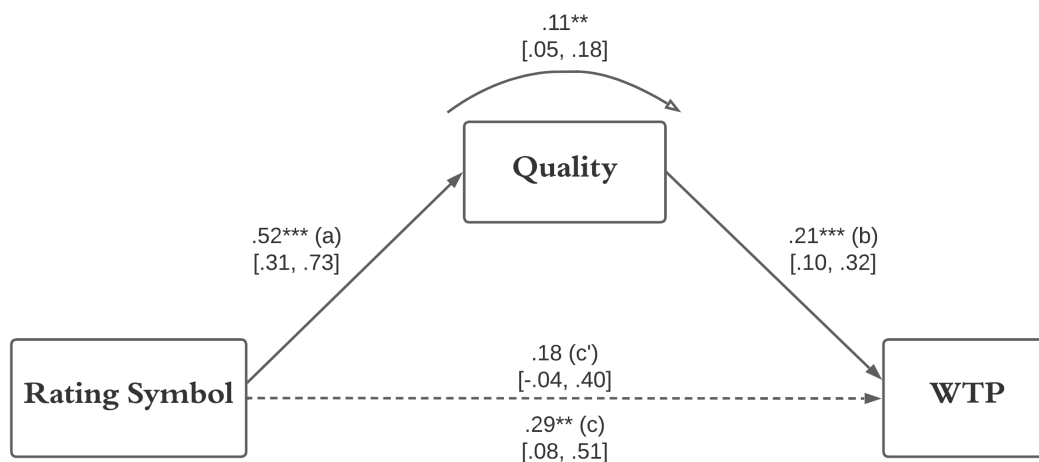
Paper Mate Profile Ballpoint Pen
Rating: 2.5 ★

How much would you be willing to pay (\$1 to \$10)?

Data Exclusion

We recruited 350 participants using the CloudResearch approved sample from the CloudResearch panel of US participants in exchange for monetary compensation. Per our pre-registered exclusion criteria, we removed any participant who met any of the following conditions: duplicate IP addresses ($n = 4$), response time to complete the survey was three standard deviations above or below the mean response time ($n = 6$), or failed to verify that the ratings were out of five stars ($n = 6$). This left us with 1336 WTP and 1336 quality judgment responses from 334 participants (47% non-male, $M_{age} = 41$ years). For WTP responses, we further excluded 21 WTP responses since these responses were more than three standard deviations above or below the mean WTP value. All quality judgment responses were within the three standard deviations above or below the mean quality judgment value. We used 1315 WTP responses and 1336 quality judgments for the analysis.

Figure W6: Mediation Analysis Results in Study 5



NOTE.—

All continuous variables are standardized. Numbers annotating curved arrows indicate indirect effects. Bracketed numbers indicate 95% CIs. * $p < .05$, ** $p < .01$, *** $p < .001$.

REFERENCES

- Bergstrom, C. T., and J. West. 2016. “*The Principle of Proportional Ink.*”
http://callingbullshit.org/tools/tools_proportional_ink.html.
- Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Graphics press
Cheshire, CT.

CHAPTER 2 APPENDIX

Contents

- 11. Supplementary Study 1
- 12. Supplementary Study 2
- 13. Study 1
- 14. Study 2
- 15. Study 3A
- 16. Study 3B
- 17. Study 4
- 18. Alternate version of Study 4
- 19. Study 5
- 20. Study 6A
- 21. Examples of companies using year and length framing optimally

SUPPLEMENTARY STUDY 1: MODERATION BY EASE OF COMPUTATION

This pre-registered study (https://aspredicted.org/YNK_XL7) has three main objectives. First, it aims to manipulate ease of computation (or processing ease) to explore how the year-length framing effect varies with time distances that differ in computational complexity within the year framing. Second, we aim to measure a process variable by assessing subjective processing ease during the subjective time perception task. Third, we seek to measure another process variable by evaluating the vividness of the length of time. Consequently, this study investigates the roles of processing ease and vividness in influencing the year-length framing effect.

Conducted in February 2024, this study follows a three-factor mixed-design approach. The first factor is a between-subjects variable with two levels (time framing: Year vs. Length). The second is a within-subjects variable with three levels (length of time distance: X , $X+10$, and $X+20$). The third factor is a between-subjects variable with three levels (type of time distance; $X = 8, 10$, or 12). Each participant rated how close or far in the future each of the three time distances (X , $X+10$, and $X+20$) feels, presented in a random order.

We used two types of time distances to manipulate computational complexity within the year framing: relatively easy computations ($X = 10$) and relatively difficult computations ($X = 8$ or 12); computations such as $2034 - 2024$ are relatively easier than computations such as $2036 - 2024$ (Thomas & Morwitz 2009). By having some participants evaluate subjective time perception for years with easier computations, we aimed to test the processing ease account by manipulating ease-of-computation. This design allowed us to examine how the year-length framing effect varies with time distances that differ in computational complexity, providing insights into the roles of processing ease in the year-length framing effect.

Method

Participants. In exchange for monetary compensation, we recruited 1628 participants using the CloudResearch MTurk Toolkit. Per our pre-registered exclusion criteria, we removed participants who met any of the following conditions: duplicate IP addresses ($n = 14$); failed fraud detection check ($n = 53$), total response time for the three subjective time distance questions ($n = 13$), or response time to complete the survey ($n = 22$) was three standard deviations above or below the mean response times. In total, we excluded 102 participants, leaving 1526 participants for the analysis (51.6% non-male, $M_{\text{age}} = 43$).

Procedure. Participants were assigned to one of six conditions defined by time framing (year vs. length) and type of time distance ($X = 8, 10$, or 12). Within each condition, participants judged how far three future time distances felt, presented in random order, similar to the methods used in Studies 4, 6A, and 6B. After completing the subjective time perception questions, participants answered both subjective processing ease questions and subjective vividness perception questions; the order of these blocks was randomized.

For the processing ease questions, participants responded to three questions in random order, each corresponding to one of the time distances they evaluated in the subjective time perception task. For example, in the $X = 10$ conditions, the processing ease questions were: “How easy or difficult was it for you to judge how far the year 2034 [10 years later] feels?”, “How easy or difficult was it for you to judge how far the year 2044 2034 [20 years later] feels?”, and “How easy or difficult was it for you to judge how far the year 2054 2034 [30 years later] feels?”, all rated on a seven-point scale (1 = Not at all easy, 7 = Very easy). This measure was adapted from Menon, Raghubir, and Schwarz (1995).

Similarly, to assess the vividness of length of time, participants were asked: “How easy is it for you to visualize the length of time from now until the year 2034 [next 10 years] when you try to imagine it?”, “How easy is it for you to visualize the length of time from now until the year 2044 [next 20 years] when you try to imagine it?”, and “How easy is it for you to visualize the length of time from now until the year 2054 [next 30 years] when you try to imagine it?”, also presented in random order and rated on a seven-point scale (1 = Not very easily imagined, 7 = Very easily imagined). This measure has been adapted from prior literature in mental imagery where its measurement is captured through self-report scales on the ease of generating imagery (Elder & Krishna, 2021; Petrova & Cialdini, 2005).

In the end, participants answered five fraud detection questions followed by standard demographic questions.

Results

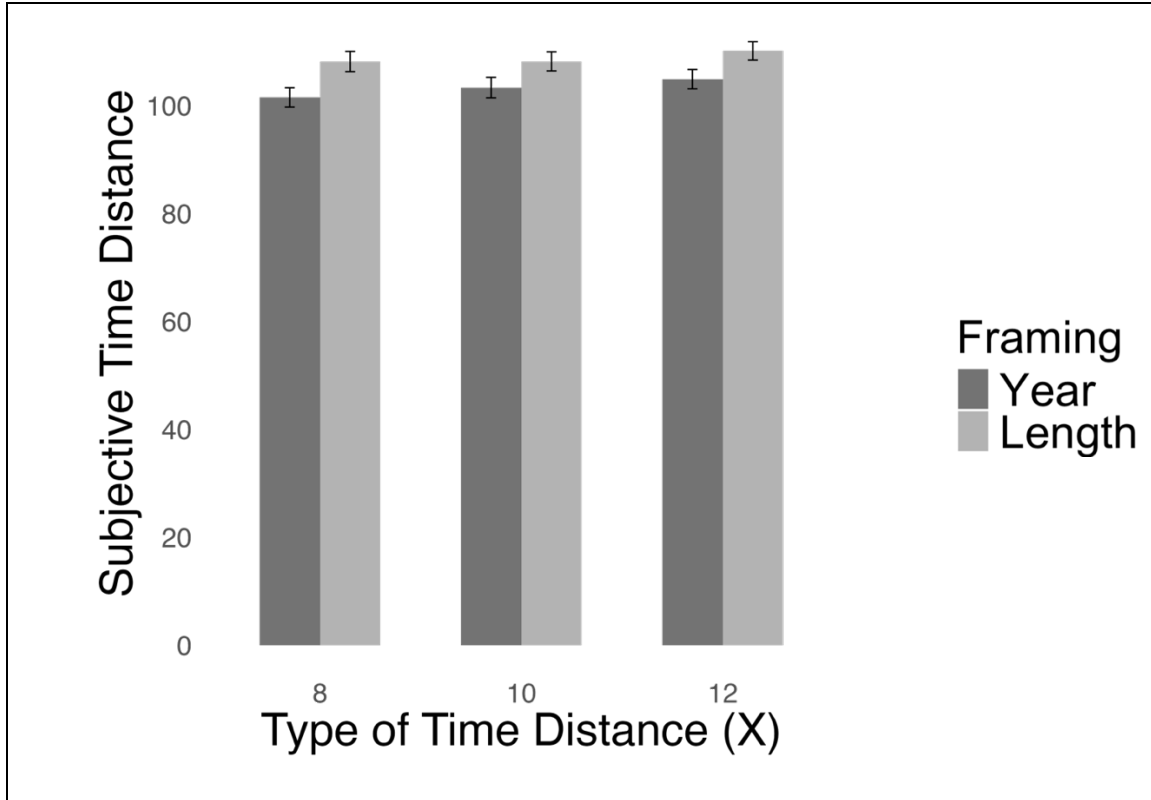
Subjective Time Perception. The subjective time perception responses were analyzed using a 2 x 3 x 3 mixed factorial ANOVA, with time framing condition (year vs. length) and type of time distance (X = 8, 10, or 12) as between-subjects factors, and time distance (X, X+10, and X+20) as a within-subjects factor. The analysis revealed significant main effects of both time framing and time distance. Specifically, participants perceived time as significantly longer in the length framing condition ($M_{\text{Length}} = 108.87$, $SD = 49.30$) compared to the year framing condition ($M_{\text{Year}} = 103.27$, $SD = 50.37$; $F(1, 1520) = 7.02$, $p = .008$, $\eta^2_p = .004$). Additionally, there was a significant effect of time distance, with perceptions of time increasing as the distance increased ($M_X = 75.60$, $M_{X+10} = 109.40$, $M_{X+20} = 133.24$; $F(2, 3040) = 3690.28$, $p < .001$, $\eta^2_p = .71$).

The main effect of type of time distance was not significant ($M_8 = 104.81$, $M_{10} = 105.83$,

$M_{12} = 107.59$; $F(2, 1520) = .55, p = .58, \eta^2_p = .000$). Importantly, the two-way interaction between time framing and type of time distance was not significant ($F(1, 813) = 8.98, p = .003, \eta^2_p = .01$). Furthermore, the time framing effect for relatively easy computations ($X = 10$) was not statistically different from the relatively difficult computation conditions ($X = 8$ or 12): $M_{10} = 4.86, M_{8_or_12} = 5.95$; $F(1, 1520) = .06, p = .81, \eta^2_p = .000$. Refer to Figure W1 for a graphical representation.

The analysis also revealed a significant interaction between time framing and time distance, indicating that the effect of time framing was stronger for longer time distances ($X+10$ and $X+20$) compared to shorter ones (X): $M_X = 1.46, M_{X+10} = 7.49, M_{X+20} = 7.81$; $F(2, 3040) = 14.08, p < .001, \eta^2_p = .01$.

Figure W1: Subjective Time Distance by Time Framing and by Type of Time Distance



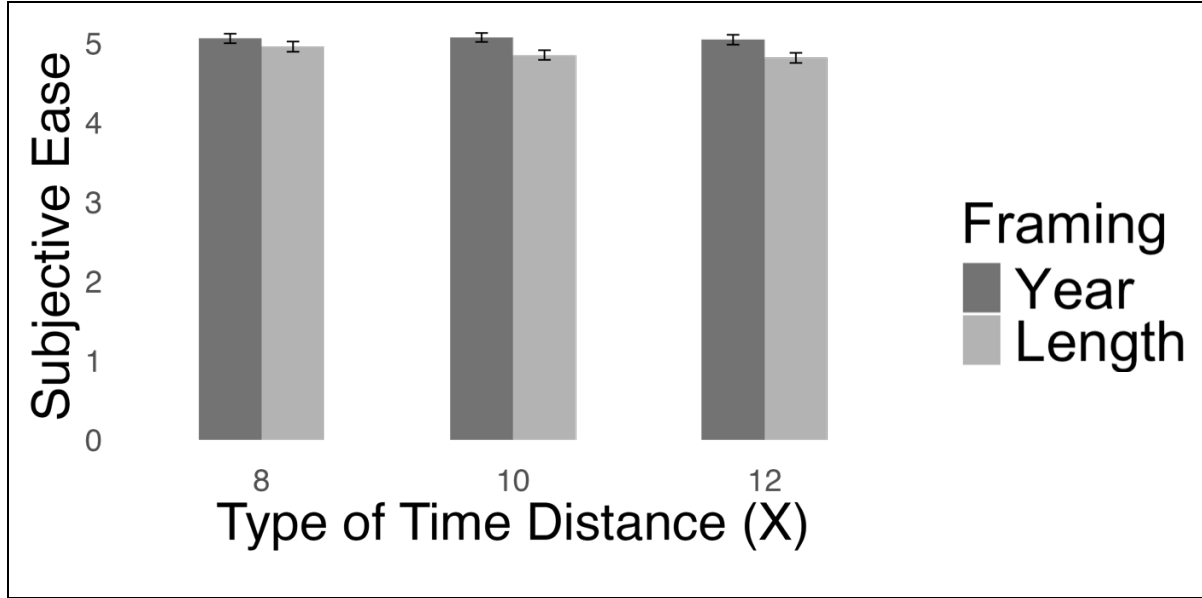
NOTE.— The figure shows raw means and bars depict standard errors.

Self-report Ease. The subjective ease responses were analyzed using the same ANOVA model. The main effect of time framing on subjective ease was significant; however, contrary to our expectations, participants rated year framing as easier than length framing ($M_{\text{Year}} = 5.06$, $M_{\text{Length}} = 4.88$; $F(1, 1520) = 6.26$, $p = .012$, $\eta^2_p = .004$). The main effect of type of time distance on subjective ease was not significant ($M_8 = 5.01$, $M_{10} = 4.96$, $M_{12} = 4.93$; $F(2, 1520) = .38$, $p = .68$, $\eta^2_p = .000$). However, the main effect of time distance was significant, indicating that longer time distances were perceived as less easy to judge ($M_X = 5.39$, $M_{X+10} = 4.87$, $M_{X+20} = 4.65$; $F(2, 3040) = 214.23$, $p < .001$, $\eta^2_p = .12$).

The two-way interaction between time framing and type of time distance was not significant ($F(2, 1520) = .30$, $p = .74$, $\eta^2_p = .000$). However, the two-way interaction between time framing and time distance approached significance ($F(2, 3040) = 2.98$, $p = .051$, $\eta^2_p = .002$).

The effect of time framing on subjective ease was weaker for shorter time distances (X) compared to longer time distances (X+10 and X+20).

Figure W2: Subjective Ease by Time Framing and by Type of Time Distance



NOTE.— The figure shows raw means and bars depict standard errors.

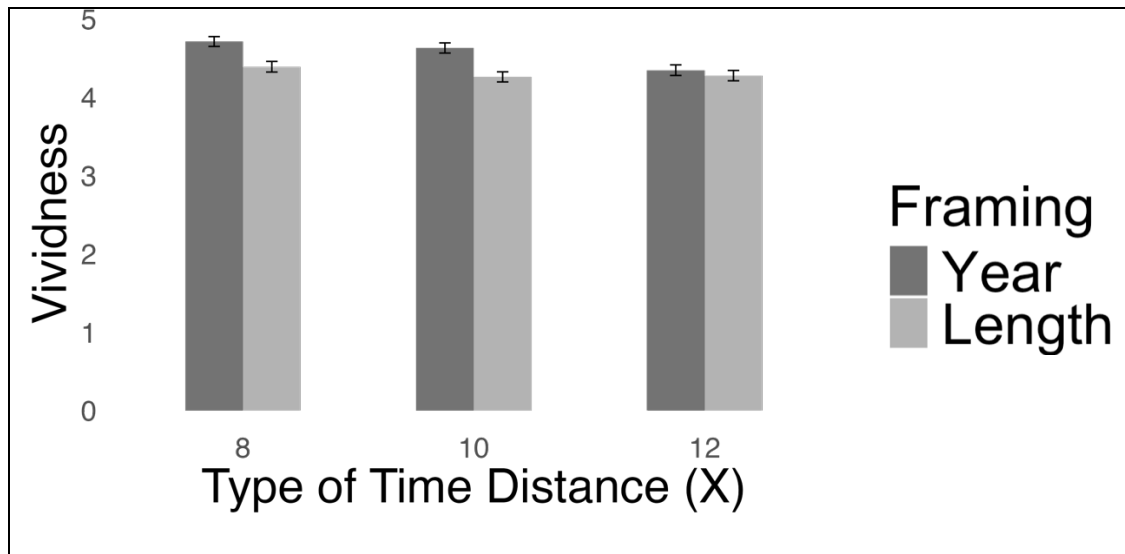
Vividness. The vividness responses were analyzed using the same ANOVA model. The main effect of time framing on vividness was significant; however, contrary to our expectations, participants rated the length of time in the year framing as more vivid than in the length framing ($M_{\text{Year}} = 4.57$, $M_{\text{Length}} = 4.31$; $F(1, 1520) = 10.40$, $p = .001$, $\eta^2_p = .006$).

The main effect of type of time distance on vividness was also significant ($M_8 = 4.56$, $M_{10} = 4.44$, $M_{12} = 4.31$; $F(2, 1520) = 3.09$, $p = .046$, $\eta^2_p = .004$). Additionally, the main effect of time distance was significant, indicating that longer time distances were perceived as less vivid ($M_X = 5.24$, $M_{X+10} = 4.32$, $M_{X+20} = 3.75$; $F(2, 3040) = 1038.40$, $p < .001$, $\eta^2_p = .41$).

The two-way interaction between time framing and type of time distance was not significant ($F(2, 1520) = 1.40$, $p = .25$, $\eta^2_p = .002$). However, the two-way interaction between time framing and time distance was significant ($F(2, 3040) = 11.73$, $p < .001$, $\eta^2_p = .008$),

indicating that the effect of time framing on vividness was weaker for shorter time distances (X) compared to longer time distances (X+10 and X+20).

Figure W3: Vividness by Time Framing and by Type of Time Distance



NOTE.— The figure shows raw means and bars depict standard errors.

Discussion

The results of Supplementary Study 1 provide several key insights. First, the time framing effect does not attenuate for time distances where the computational complexity for the year framing is lower. This finding suggests that the observed effect of time framing on subjective time perception is not driven by the ease of computation.

Second, our analyses revealed that both subjective ease and vividness were significantly higher in the year framing than in the length framing. Therefore, these two accounts—ease of computation and vividness—can be ruled out as primary drivers of the year-length framing effect, as these accounts assumed that ease of computation and vividness are higher in the length (vs. year) framing.

Overall, these results support our hypothesis that the year-length framing effect is

influenced by the perception of numerical change rather than by computational ease or vividness. This reinforces the idea that larger nominal values in the year framing lead to a smaller perceived change, consistent with the logarithmic nature of the mental number line.

SUPPLEMENTARY STUDY 2: TWO FUTURE TIME POINTS

Supplementary Study 2 had two aims. First to test the generalizability of year-length effect with time intervals not including the present (e.g., comparing intervals with year framing—2032 to 2043—with equivalent time intervals described using length framing—11 years). The second aim was to test whether the year-length effect attenuates for time intervals where the ease of computation in the year framing is higher (i.e., relative computational ease between year and length framing is lower).

Conducted in March 2024 ($N = 155$), this study follows a two-factor between-subjects design. The first factor is a between-subjects variable with two levels (time framing: Year vs. Length). The second is also a between-subjects variable with two levels (time distance: 10 vs. 11). Thus, subjects were assigned to one of the four conditions; see images of the question asked in each of the four conditions.

Year – 10

A	B
2030	2040

A and B are two time points in the future.

Please indicate how long the duration between A and B feels to you.

feels
very short

feels
very long

Year – 11

A	B
2032	2043

A and B are two time points in the future.

Please indicate how long the duration between A and B feels to you.

feels
very short

feels
very long

Length – 10

A B

—10 years—

A and B are two time points in the future.

Please indicate how long the duration between A and B feels to you.

feels
very short

feels
very long

→

Length – 11

A B

—11 years—

A and B are two time points in the future.

Please indicate how long the duration between A and B feels to you.

feels
very short

feels
very long

→

Results

Subjective Time Perception. The subjective time perception responses were analyzed using a 2 x 2 ANOVA, with time framing condition (year vs. length) and time distance (10 or 11) as between-subjects factors. The analysis revealed significant main effect of time framing. Specifically, participants perceived time as significantly longer in the length framing condition ($M_{\text{Length}} = 110.5$, $SD = 40.95$) compared to the year framing condition ($M_{\text{Year}} = 97.28$, $SD =$

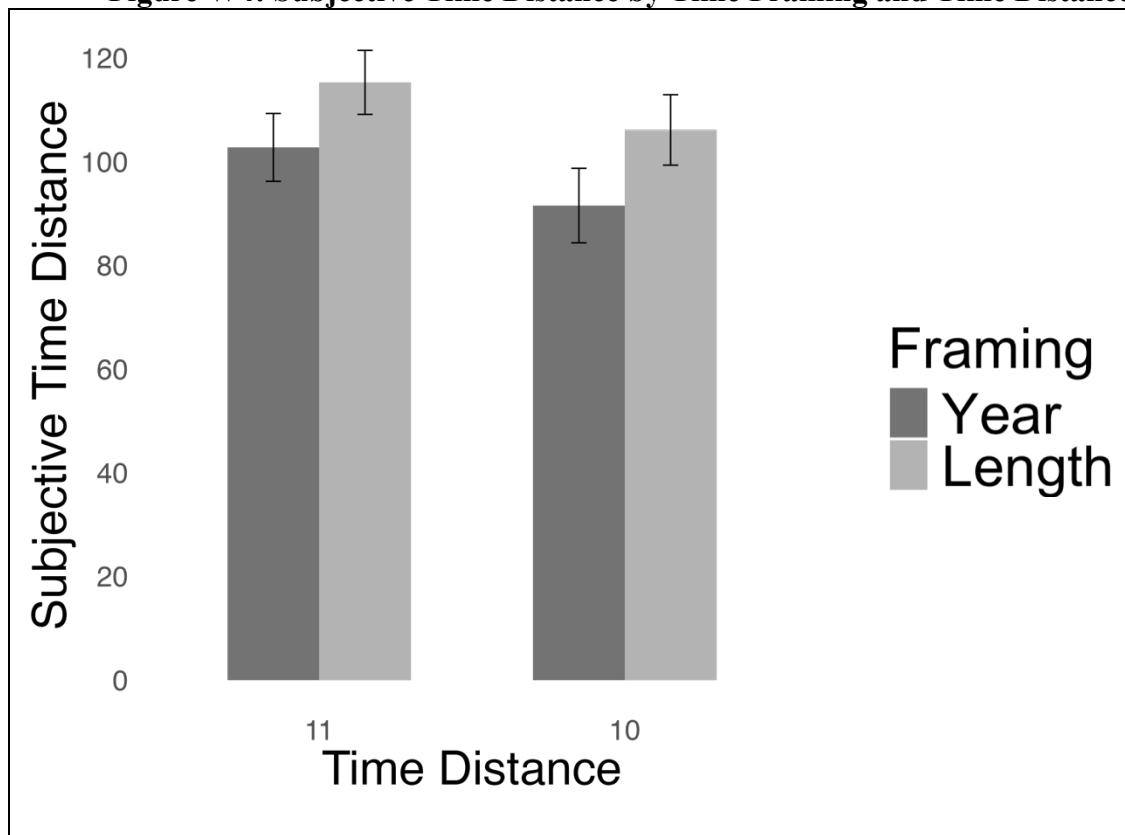
42.37; $F(1, 151) = 4.11, p = .044, \eta^2_p = .03$). Additionally, the main effect of time distance was shy of significance probably due to lower sample size ($M_{10} = 99.19, M_{11} = 108.93; F(1, 151) = 2.32, p = .13, \eta^2_p = .02$).

Importantly, the two-way interaction between time framing and time distance was not significant ($F(1, 151) = .02, p = .88, \eta^2_p = .00$). Refer to Figure W4 for a graphical representation.

Discussion

These results show that year-length effect generalizes to time intervals not including the present and that higher computational ease in the year framing does not attenuate the year-length effect.

Figure W4: Subjective Time Distance by Time Framing and Time Distance



NOTE.— The figure shows raw means and bars depict standard errors.

STUDY 1

1. Examples of length and year framed whiskey bottles

Length Framed Whiskey

Ardbeg - 30 Year Old (Very Old)



Year Framed Whiskey

Ardbeg - 1975 Vintage



2. *Examples of whiskey bottles that include both length and year framing in the name of the whiskey, but one framing is more prominent on the bottle. Length framing is more prominent in the left picture, while year framing is more salient in the right picture.*

Macallan - 26 Year Old (1993 The Scottish Gantry)



Macallan - 25 Year Old (1992) Carn Mor



3. *Descriptive statistics by top 10 brands (in frequency) and top 7 type of whiskeys*

<i>Brand</i>	<i>Price</i>	<i>Age</i>	<i>Size (in cl)</i>	<i>N</i>	<i>Type of Whiskey</i>	<i>Price</i>	<i>Age</i>	<i>Size (in cl)</i>	<i>N</i>
Macallan	310	14.2	65	3875	Single Malt	£189	15.0	69	39593
Glenfiddich	187	15.8	70	1903	Blended	£166	16.4	69	5135
Highland Park	153	18.3	70	1833	Bourbon	£304	12.5	75	3720
Balvenie	247	16.9	76	1422	Rye	£254	10.0	74	744
Glenfarclas	120	14.3	70	1410	Rum	£108	13.9	70	493
Hibiki	118	14.0	71	1255	Pure Malt	£269	16.1	70	221
Glenlivet	492	17.8	64	1115	Irish Whiskey	£178	15.1	70	191
Glendronach	250	13.1	69	1096	All types	£195	14.9	70	50916
Ardbeg	91	15.6	71	1084					
Glenmorangie	163	19.8	69	1043					
All Brands	£195	14.9	70	50916					

4. ***Robustness check 1:*** *For this robustness check, we focused exclusively on Single Malt Whiskeys, as this was the most frequent type in our dataset. By analyzing only Single Malt*

Whiskeys, we aimed to determine whether the observed effects held true within this specific category.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£271	15.8	241	6845
Length	£172	14.8	524	32748
Total	£189	15	616	39593

Robustness Check 1: OLS Regression Model for Single Malt Whiskeys

Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.099*** (-.12, -.08)
Age	.513*** (.51, .52)
Size	.149*** (.14, .16)
Strength	.359*** (.35, .37)
Delivery weight	.145*** (.14, .15)
VAT applicable [€] : Yes	-.055** (-.08, -.03)
Rarity descriptor: Very Rare [€]	.541*** (.48, .61)
Rarity descriptor: Rare [€]	.394*** (.36, .43)
Constant	3.078*** (1.50, 4.65)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes
Warning labels [€] ?	Yes
Award count?	Yes
Description length?	Yes
Bottle number and Lot no band [€]	Yes
Year of manufacture [€] ?	No
Observations	39,593
Adjusted R ²	.716
F Statistic	44.33***
DF	2298; 37294

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

5. **Robustness check 2:** For this robustness check, we included only those whiskey brands that demonstrated some variation in their time framing. Specifically, we selected brands where more than 10% of their auctions used "year" framing and more than 10% used "length" framing. By applying these criteria, we excluded brands that predominantly used a single time framing measure, thereby ensuring a more balanced and comprehensive analysis. We found 133 such brands in our dataset, allowing us to focus on those with a diverse representation of both time framing types, providing a more robust examination of the data. 27% of auctions used using year framing in this dataset.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£225	16.3	133	5060
Length	£175	14.3	133	13842
Total	£189	14.9	133	18902

Robustness Check 2: OLS Regression Model; Brands with Time Framing Variation

Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.098*** (-.12, -.08)
Age	.427*** (.42, .44)
Size	.180*** (.17, .19)
Strength	.377*** (.37, .39)
Delivery weight	.142*** (.13, .15)
VAT applicable [€] : Yes	-.008 (-.06, .04)
Rarity descriptor: Very Rare [€]	.529** (.45, .61)
Rarity descriptor: Rare [€]	.418*** (.36, .48)
Constant	5.390*** (3.54, 7.28)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes

Warning labels [€]	Yes
Award count?	Yes
Description length?	Yes
Bottle number and Lot no band [€]	Yes
Year of manufacture [€]	No
Observations	18,902
Adjusted R ²	.680
F Statistic	23.006***
DF	1822; 17079

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

6. **Robustness check 3:** For this robustness check, we included only those whiskey brands that use year framing more than a certain threshold. Specifically, we selected brands where more than 20% of their auctions used year framing. By applying this criterion, we excluded brands that predominantly used length framing, thereby addressing the norm-based explanation. We identified 370 such brands totaling 11,029 auctions in our dataset, allowing us to focus on those with a diverse representation of both time framing types and providing a more robust examination of the data. **In this subset, about 50% of auctions used year framing,** ensuring a balanced analysis and further validating our findings.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£236	14.5	370	5495
Length	£156	15.3	108	5534
Total	£196	14.9	370	11029

Robustness Check 3: OLS Regression Model;

Brands with at least 20% whiskeys described using Year framing

Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.066*** (-.09, -.04)
Age	.353*** (.34, .36)
Size	.124*** (.11, .14)
Strength	.260*** (.25, .27)
Delivery weight	.148*** (.14, .15)
VAT applicable [€] : Yes	-.096*** (-.15, -.04)

Rarity descriptor: Very Rare [€]	.820*** (.69, .95)
Rarity descriptor: Rare [€]	.078* (.00, .15)
Constant	3.720*** (2.67, 4.77)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes
Warning labels [€] ?	Yes
Award count?	Yes
Description length?	Yes
Bottle number and Lot no band [€]	Yes
Year of manufacture [€] ?	No
Observations	11,029
Adjusted R ²	.777
F Statistic	20.046***
DF	2015; 9013

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

7. **Robustness check 4:** For this robustness check, we analyzed only the most frequent brand in our dataset, Macallan. By concentrating on Macallan, we aimed to determine whether the observed effects were consistent within this specific brand. This approach allowed us to control for variations between different brands and provided a clearer understanding of the effects within the Macallan brand. This focused analysis helped us verify the robustness of our findings in the context of a single, highly represented brand.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£659	19.2	1	602
Length	£245	13.3	1	3273
Total	£309	14.2	1	3875

Robustness Check 4: OLS Regression Model for Macallan Brand

Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.425*** (-.51, -.34)
Age	.339* (.07, .61)
Size	.21*** (.19, .22)
Strength	.619*** (.58, .66)
Delivery weight	.160*** (.14, .18)
VAT applicable [€] : Yes	.012 (-.11, .14)
Rarity descriptor: Very Rare [€]	.478 (-.02, .97)
Rarity descriptor: Rare [€]	.575*** (.41, .75)
Constant	2.145** (.60, 3.69)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes
Warning labels [€] ?	Yes
Award count?	Yes
Description length?	Yes
Bottle number and Lot no band [€]	Yes
Year of manufacture [€] ?	No
Observations	3,875
Adjusted R ²	.800
F Statistic	14.07***
DF	1187; 2687

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

8. **Robustness check 5:** For this robustness check, we focused on analyzing the data within a log-transformed price range centered around the mean log price plus or minus one standard deviation of the log price. Given that the overall price range in our dataset spans from £7 to £2325, this criterion defined a new dataset with prices ranging between £40 and £290. This approach allowed us to test whether the observed effects held true within a price range closer to the mean, providing a more focused analysis of the relationship between price and other variables in our dataset.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£124	15.1	364	5788
Length	£111	14.6	794	28384
Total	£113	14.7	987	34172

Robustness Check 5: OLS Regression Model for Prices Between £40 and £290

Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.038*** (-.05, -.02)
Age	.295*** (.29, .30)
Size	.039*** (.03, .05)
Strength	.194*** (.19, .20)
Delivery weight	.119*** (.11, .13)
VAT applicable [€] : Yes	-.012 (-.04, .01)
Rarity descriptor: Very Rare [€]	.466*** (.41, .52)
Rarity descriptor: Rare [€]	.185*** (.16, .21)
Constant	5.006*** (3.77, 6.24)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes
Warning labels [€] ?	Yes
Award count?	Yes
Description length?	Yes
Bottle number and Lot no band [€]	Yes
Year of manufacture [€] ?	No
Observations	34,172
Adjusted R ²	.473
F Statistic	12.40***
DF	2692; 31479

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

9. **Robustness check 6:** For this robustness check, we excluded whiskeys aged less than 5 years from our analysis. Auctioned whiskeys are typically older and have more established market values and reputations. The range of ages in our dataset spans from 1 to 30 years, and by focusing on whiskeys aged 5 years or older, we ensure that we are analyzing more mature and typically more sought-after products. This approach allows us to better capture the characteristics and market dynamics of whiskeys that are more representative of those commonly found in auctions, providing a clearer understanding of their value and appeal.

<i>Framing</i>	<i>Price</i>	<i>Age</i>	<i>Unique brands</i>	<i>N</i>
Year	£270	16.3	443	7500
Length	£183	14.9	1065	42616
Total	£113	14.7	987	34172

Robustness Check 6: OLS Regression Model for Whiskeys Aged 5 Years or More

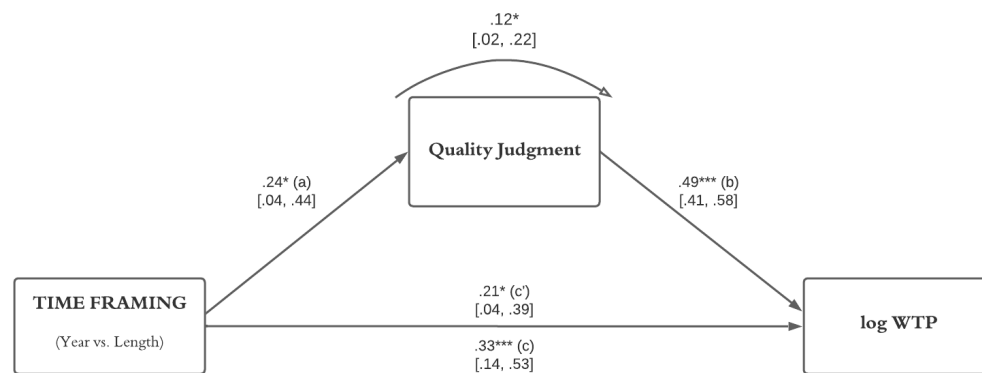
Predictor Variables	Dependent Variable: (log) Price
Framing [€] : Year	-.100** (-.12, -.08)
Age	.505*** (.50, .51)
Size	.142*** (.14, .15)
Strength	.354*** (.35, .36)
Delivery weight	.148*** (.14, .15)
VAT applicable [€] : Yes	-.033** (-.06, -.01)
Rarity descriptor: Very Rare [€]	.570*** (.51, .63)
Rarity descriptor: Rare [€]	.339*** (.31, .37)
Constant	2.080** (.54, 3.62)
Brand [€] ?	Yes
Year of sale [€] ?	Yes
Type of whiskey [€] and Region [€] ?	Yes
Distillery status [€] ?	Yes
Warning labels [€] ?	Yes
Award count?	Yes
Description length?	Yes

Bottle number and Lot no band [€]	Yes
Year of manufacture [€] ?	No
Observations	50,116
Adjusted R ²	.737
F Statistic	47.38***
DF	3023; 47092

NOTE.—All continuous independent variables are standardized. [€] indicates categorical variables. Bracketed numbers indicate 95% CIs, * $p < .05$, ** $p < .01$, *** $p < .001$.

STUDY 2

1. Mediation model



NOTE.— All continuous variables are standardized. Numbers annotating curved arrows indicate indirect effects. Dashed lines show statistically not significant paths. Bracketed numbers indicate 95% CIs. * $p < .05$, ** $p < .01$, *** $p < .001$, ^ $p < .05$, exact value N/A.

Study 3A

2. ANCOVA analysis for subjective investment return

<i>Model Term</i>	<i>df1</i>	<i>df2</i>	<i>F ratio</i>	<i>p value</i>
Time_Framing	1	428	8.20	.004
Time_Distance	3	1302	122.83	<.0001
House_Ownership	1	428	.03	.866
Invest_in_Stocks	1	428	.60	.440
Age*	1	428	15.05	<.001
(log) Income	1	428	.50	.482
Gender	1	428	1.63	.202
Education	1	428	.01	.931
Time_Framing x Time_Distance	3	1302	.80	.494

*Older people judged the investment returns as more favorable.

Study 3B

1. Exact information given to participants about CDs

“A certificate of deposit (CD) is a product offered by banks and credit unions that provides an interest rate premium in exchange for the customer agreeing to leave a lump-sum deposit untouched for a predetermined period of time.”

2. ANCOVA analysis for subjective investment return

<i>Model Term</i>	<i>df1</i>	<i>df2</i>	<i>F ratio</i>	<i>p value</i>
Time_Framing	1	426	9.77	.002
Time_Distance	3	1293	221.34	<.0001
CD_Investment_Experience*	1	428	9.45	.002
Age*	1	428	6.74	.010
(log) Income	1	428	1.01	.32
Gender	1	428	.36	.55
Education*	1	428	4.26	.04
Time_Framing x Time_Distance	3	1293	3.17	.02

* Participants who have prior experience with CD investments, possess higher levels of education, and are older people perceived the returns on CD investments as more favorable.

Study 4

1. Subjective Time Perception Task (length condition)

Participants responded by moving the cursor towards the right or left.

How far does **20 years ago** feel?

feels
very recent

feels
very distant



2. Subjective Time Perception Task (year condition)

How far does the **year 2002** feel?

feels
very recent

feels
very distant



3. Descriptive statistics

	Time Framing	
	Length	Year
Time	3 years	2019
Subjective Time Distance	86	67
Time	6 years	2016
Subjective Time Distance	117	100
Time	20 years	2002
Subjective Time Distance	157	147

4. ANCOVA analysis for subjective time perception

Model Term	df1	df2	F ratio	p value
Time_Framing	1	614	25.06	<.0001
Time_Distance	2	1236	932.19	<.0001
Age*	1	614	4.10	.043

(log) Income	1	614	1.56	.21
Gender	1	614	1.91	.17
Education*	1	614	.13	.72
Time_Framing x Time_Distance	2	1236	3.55	.03

**Older people judged subjective time distances as shorter.*

Alternate Version of Study 4

1. Subjective Time Perception Task (length condition)

The starting point of the cursor was hidden.

How far does **2 years** later feel?

feels
very close

feels
very far



2. Subjective Time Perception Task (year condition)

How far does same time of the **year 2025** feel?

feels
very close

feels
very far



3. Subjective Time Perception Task (both length and year condition)

same time of the **year 2025**
2 years later

How far does this time feel?

feels
 very close

feels
 very far



4. Descriptive statistics

	Time Framing		
	Length	Year	Both
Time	2 years	2024	2024 (2 years)
Subjective Time Distance	73.6	38.5	38.4
Time	5 years	2027	2027 (5 years)
Subjective Time Distance	112.1	81.2	79.1
Time	7 years	2029	2029 (7 years)
Subjective Time Distance	131.3	98.3	98.8
Time	10 years	2032	2032 (10 years)
Subjective Time Distance	150.2	125.2	122.9

5. ANCOVA analysis for subjective time perception

Model Term	df1	df2	F ratio	p value
Time_Framing	2	401	5.20	.006
Time_Distance	3	1215	1033.90	<.0001
Age*	1	401	3.91	.049
(log) Income	1	401	.66	.42
Gender	1	401	2.31	.13
Education	1	401	1.69	.19
Time_Framing x Time_Distance	6	1215	1.83	.09

*Older people judged subjective time distances as shorter.

Study 5

1. Prompt for intertemporal choice task

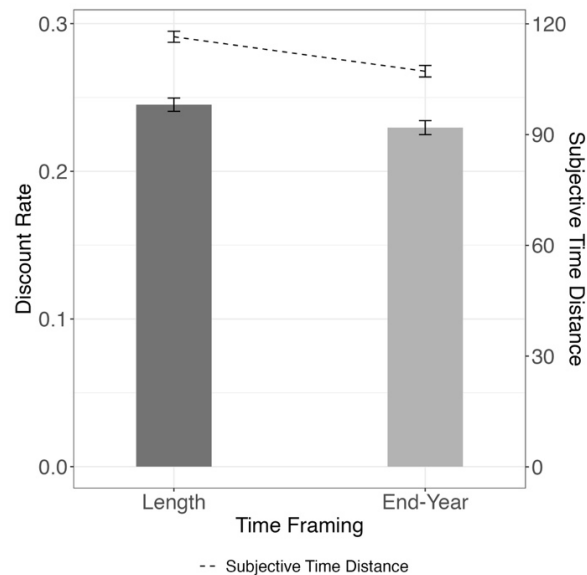
*“Imagine the state you live in has a budget surplus that it is planning to pay out as rebates of \$1000 to **each** citizen. Alternatively, the state is also considering investing the surplus in endowment funds that will yield a **higher amount in the future**. The investment would allow the state to offer a higher rebate amount, to be paid in the future.*

The state wants to know whether you would prefer a \$1000 rebate immediately, or a higher rebate amount in the future. You will be able to avail the future rebates even if you move away from your current state. If you prefer a higher amount in the future, the state wants to know what amount would be sufficient for you to delay your rebate.”

2. ‘money now’ and ‘money later’ values used for different time distances

Today	in 2 years [2024] on the same day	in 3 years [2025] on the same day	in 4 years [2026] on the same day
\$1000	\$1100	\$1100	\$1100
\$1000	\$1200	\$1300	\$1400
\$1000	\$1300	\$1500	\$1700
\$1000	\$1500	\$1700	\$2000
\$1000	\$1600	\$1900	\$2300
\$1000	\$1700	\$2100	\$2600
\$1000	\$2100	\$2500	\$3000

3. Figure W5: Discount rate and subjective time perception



NOTE.— The figure shows raw means from Study 5 and bars depict standard errors.

4. Descriptive Statistics

	Time Framing	
	Length	Year
Time	2 years	2024
Discount Rate	.268	.237
Subjective Time Distance	108.76	97.83
Time	3 years	2025
Discount Rate	.232	.223
Subjective Time Distance	117.25	103.68
Time	4 years	2026
Discount Rate	.236	.228
Subjective Time Distance	123.16	120.66

5. ANCOVA analysis for discount rate

<i>Model Term</i>	<i>df1</i>	<i>df2</i>	<i>F ratio</i>	<i>p value</i>
Time_Framing	1	1679	5.86	.016
Time_Distance	2	1679	5.67	.004
Age*	1	1679	12.67	<.001
(log) Income*	1	1679	12.12	<.001
Gender*	1	1679	14.54	<.001
Education*	1	1679	28.10	<.0001
Time_Framing x Time_Distance	2	1679	1.04	.35

*Older people, people with higher income, males, and people with more education exhibited lower discount rates.

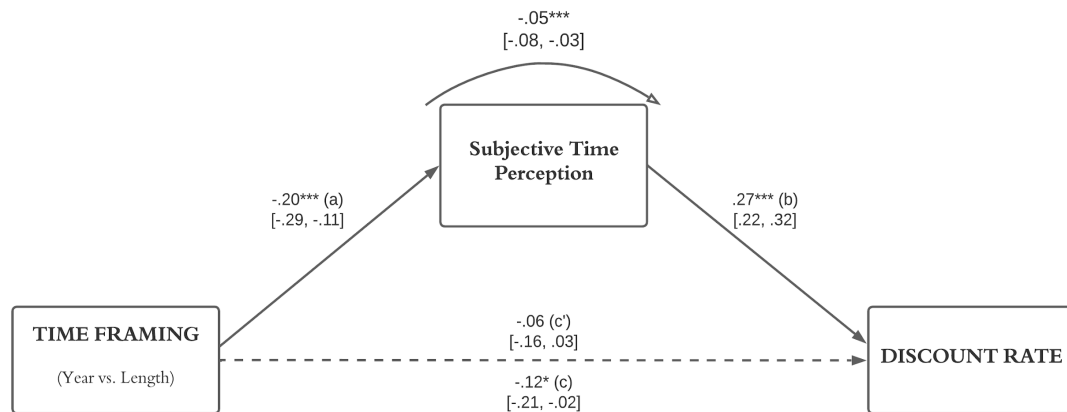
6. ANCOVA analysis for subjective time perception

<i>Model Term</i>	<i>df1</i>	<i>df2</i>	<i>F ratio</i>	<i>p value</i>
Time_Framing	1	1679	17.81	<.0001
Time_Distance	2	1679	26.28	<.0001
Age*	1	1679	23.48	<.0001
(log) Income*	1	1679	12.60	<.001
Gender	1	1679	2.28	.132
Education	1	1679	1.11	.293
Time_Framing x Time_Distance	2	1679	2.42	.090

*Older people and people with higher income judged subjective time distances as shorter.

7. Mediation analysis

Figure W6: Mediation Model in Study 5



NOTE.— All continuous variables are standardized. Numbers annotating curved arrows indicate indirect effects. Dashed lines show statistically not significant paths. Bracketed numbers indicate 95% CIs. * $p < .05$, ** $p < .01$, *** $p < .001$, ^ $p < .05$, exact value N/A.

Study 6A

1. Due to a technical error, we did not receive responses from majority of participants for four questions in this study: age, gender, income, and education.
2. Fraud detection questions used in Study 6A and 6B. Participants who did not answer ‘no or none of the above’ for two or more questions were excluded from the analysis.

Have you seen any live music shows in Shea Stadium, NY in the last 2 years?

Yes

No

In future studies, we may target individuals who have recently taken a cruise on the following cruise ships. Please indicate if you have taken a cruise with any of the following companies in the last 2 years:

Sail Through Vacations

Yacht-ify Cruises

Pacific Travails LTD

Outstanding Ocean Outings

None of the above

Have you filed a homeowners insurance claim due to damage from lightning within the past 3 months?

Yes

No

Please select which of the following hair-care products you have used in the last 6 months:

Bunfocore

Pleistorene Grow

Truefolica Treatment

None of the above

Have you recently purchased food from any of these small food chain stores?

Cheese & Wine Flavor Junction

Zesty's Tomato Pies

Paulina's Poutine King

Toasty Pita BBQ

None of the above

3. *Numerical change perception questions in year framing, shown one at a time and in random order.*

How big does the change from **1000 to 1005** feel to you?

very small

very big



How big does the change from **1000 to 1009** feel to you?

very small

very big



4. *Numerical change perception questions in length framing, shown one at a time and in random order.*

How far does **5 years** later feel from now?

feels
very close

feels
very far



How far does **9 years** later feel from now?

feels
very close

feels
very far



Examples of companies using year and length framing optimally

We found that it is quite common for companies to highlight the number of years they have been in operation to signal a long history. While it may be impractical to etch the number of years into a permanent facade, as this would require updating every year, marketing campaigns and digital press releases are not permanent and can be updated regularly. Numerous examples on LinkedIn and other platforms show company campaigns using length framing to signal their rich and long history. See below for examples from two brands: Honda and Dell.

Excerpt from a Honda Press Release in 2023; Source: [link](#)

September 29, 2023 – TORRANCE, Calif.

- New "Keep Dreaming" brand spot showcases Honda's commitment to achieving global carbon neutrality by 2050
- Campaign introduces first all-electric Honda SUV, the 2024 Prologue, along with past Honda breakthroughs and future products and innovations
- "Wall of Dreams" epic, interactive brand mural and AR experience in New York and Los Angeles features Honda dreams that have inspired and moved people globally for 75 years

About Honda

Honda offers a full line of clean, safe, fun and connected vehicles sold through more than 1,000 independent U.S. Honda dealers. The award-winning Honda lineup includes the Civic and Accord, along with the HR-V, CR-V, Passport and Pilot sport utility vehicles, the Ridgeline pickup and the Odyssey minivan. Honda's electrified vehicle lineup includes the Accord hybrid, CR-V hybrid, and, in the future, Civic hybrid. The Honda Prologue SUV, Honda's first volume battery-electric vehicle, will join the lineup in 2024.

Honda has been producing automobiles in America for over 40 years and currently operates 18 major manufacturing facilities in North America. In 2022, more than 99% of all Honda vehicles sold in the U.S. were made in North America, with more than two-thirds made in America, using domestic and globally sourced parts.

More information about Honda is available in the [Digital FactBook](#).

In the provided press release, Honda strategically uses different framing techniques based on their goals. When signaling achievements to be realized soon, they use year framing, such as "achieving global carbon neutrality by 2050," emphasizing the short span of time until the goal is reached. Conversely, when highlighting their long history and sustained efforts, they use length framing, such as "producing automobiles in America for over 40 years" and "moved people globally for 75 years," to convey the extensive duration of their contributions and legacy. These framing strategies allow Honda to effectively emphasize different lengths of time, with year framing highlighting shorter future intervals and length framing underscoring long-term achievements and historical milestones. Additionally, we found that Honda has been producing automobiles in America since 1979 (which makes it 44 years until 2023), so using "for over 40 years" instead of "since 1979" is a striking example consistent with our recommendation.

Dell LinkedIn Post; Source: [link](#)



Dell Technologies

5,012,124 followers

22h • 🌐

[+ Follow](#)

We're celebrating 40 years of innovation, and we're going big. Check us out on one of the coolest LED screens in the world. 😎 [#DellTechWorld](#)
[#SphereVegas](#) [Sphere](#)

