Fooled by Stars:

Perceptual Biases in Judgments of Numeric Ratings

Deepak Sirwani

Srishti Kumar

Manoj Thomas

Data, Code, Materials: https://bit.ly/3YgBJs1

This version: February 20, 2023

Please do not share without permission.

Deepak Sirwani (ds2344@cornell.edu) is a PhD candidate in marketing at the SC Johnson College of Business at Cornell University, Sage Hall, 114 E Ave, Ithaca, NY 14853. **Srishti Kumar** (phd18srishtik@iima.ac.in) is a PhD student in Marketing at the Indian Institute of Management, Ahmedabad, India. **Manoj Thomas** (manojthomas@cornell.edu) is the Nakashimato Professor of Marketing at the SC Johnson College of Business at Cornell University. All correspondence regarding this manuscript should be addressed to Deepak Sirwani.

Abstract

Numerical ratings are frequently used to inform evaluative judgments of products and services. This research shows that the type of perceptual symbol used to communicate ratings can bias people's evaluative judgments. People tend to overestimate the magnitude of ratings when graphical symbols are used (e.g., image of three and a half stars) and underestimate the magnitudes when Arabic numerals are used (e.g., 3.5). These biases are only observed for fractional ratings, not for round ratings. The overestimation bias in graphic ratings is caused by the visual completion of incomplete images, leading people to anchor on rounded-up numbers. In contrast, the underestimation bias in Arabic numeral ratings is caused by left-digit anchoring, leading people to anchor on rounded-down numbers. As a result, retailers who use stars or circles for ratings may have an unfair advantage, as their ratings might be perceived to be higher than they are. Conversely, retailers using Arabic numeral ratings may be at a disadvantage, as their ratings may be underestimated. Our findings highlight the significance of perceptual processes in numerical cognition and demonstrate that the type of perceptual symbol used to communicate ratings can materially influence consumers' quality perceptions and willingness to pay.

(196/200 words)

Keywords:
ratings; magnitude judgment; perceptual bias; visual completion; left-digit anchoring; numerical cognition

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, 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. 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), while others, such as Uber, Facebook, and Airbnb, use Arabic numerals (e.g., 3.5). Some, such as Google Maps, Walmart, and Goodreads, use a combination of both types of symbols—stars and numerals together—to represent ratings. In this research we examine how the format of the ratings influences evaluative judgments. Do people's 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?

We report results from a series of experiments 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). We find that people tend to overestimate the magnitude of star ratings and underestimate that of Arabic numeral ratings. Furthermore, the bias in the perception of ratings is only observed for fractional ratings (e.g., 3.2, 3.5, 3.7, etc.), not for round ratings (e.g., 2.0, 3.0, 4.0, etc.). Our experiments were designed to characterize the perceptual and cognitive processes that underlie these biases in magnitude judgments of ratings. Our findings show that, depending on the type of symbolic representation used, different types of perceptual biases influence the transcoding of symbols to subjective magnitude judgments. When graphic symbols are used, the transcoding is biased by visual completion that results in overestimation of ratings. When Arabic numerals are used, the transcoding is biased by left-digit anchoring that results in underestimation of ratings.

Our results have obvious implications for managers, suggesting that retail managers should pay attention to the type of rating symbols they use. Additionally, our findings also have implications for theory, intersecting and contributing to three distinct research streams in consumer behavior that have been evolving in parallel: visual perception, product ratings, and numerical cognition. Although there is a rich stream of research documenting how perceptual representations, perceptual organization, and perceptual processes influence downstream variables (Bagchi and Cheema 2013; Barasz et al. 2017; Hagtvedt and Brasel 2017; Krishna 2006; Raghubir and Krishna 1999; Townsend and Kahn 2014), the literature on product ratings and numerical cognition have largely ignored the role of perceptual symbols. 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). This is the first research to study how merely changing perceptual symbols without changing the information content—using stars instead of Arabic numerals to represent the same ratings—can influence judgments.

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; Pandelaere, Briers, and Lembregts 2011; Sevilla, Isaac, and Bagchi 2018) 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 in 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 perceptual symbols in numerical cognition and product ratings. We present a conceptual framework that can be used to explain how people encode the magnitude of fractional numbers, predicting when fractional numbers will be overestimated or underestimated, and the role of perceptual symbols in such biases.

The remainder of this article is organized as follows. We begin by reviewing prior research on numerical cognition and visual perception that lay the foundation for our predictions about perceptual biases in magnitude judgments of numerical quantities. Then we present results from studies that test these predictions.

THEORETICAL BACKGROUND

Our conceptualization of the impact of symbols on numerical magnitude judgments is based on two principles. First, processing fractional numbers can be challenging (relative to whole numbers), causing people to rely on heuristics for magnitude judgments. Specifically, we propose that people use whole numbers as anchors to judge the magnitude of fractional numbers. Second, perceptual biases can play a role in determining these anchors and may vary based on the type of symbol used. In the case of graphical symbols, visual completion bias affects the anchor, whereas for Arabic numerals, the left-digit bias shapes it.

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 transcoding of symbolic representation to analog magnitudes (Dehaene 1997; Dehaene, Dupoux, and Mehler 1990; Hinrichs, Yurko, and Hu 1981). Inaccurate transcoding 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. However, this is not true for fractions. Children are able to count from a young age, but they do not understand fractions until later in their development (Dehaene 1997). When encountering fractional numbers, such as 3.2, 3.5, or 3.8, people may struggle to instinctively assess their magnitudes (Gigerenzer and Hoffrage 1999). In these situations, 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). 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, this adjustment process is 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 anchor chosen was a rounded-up or a rounded-down number.

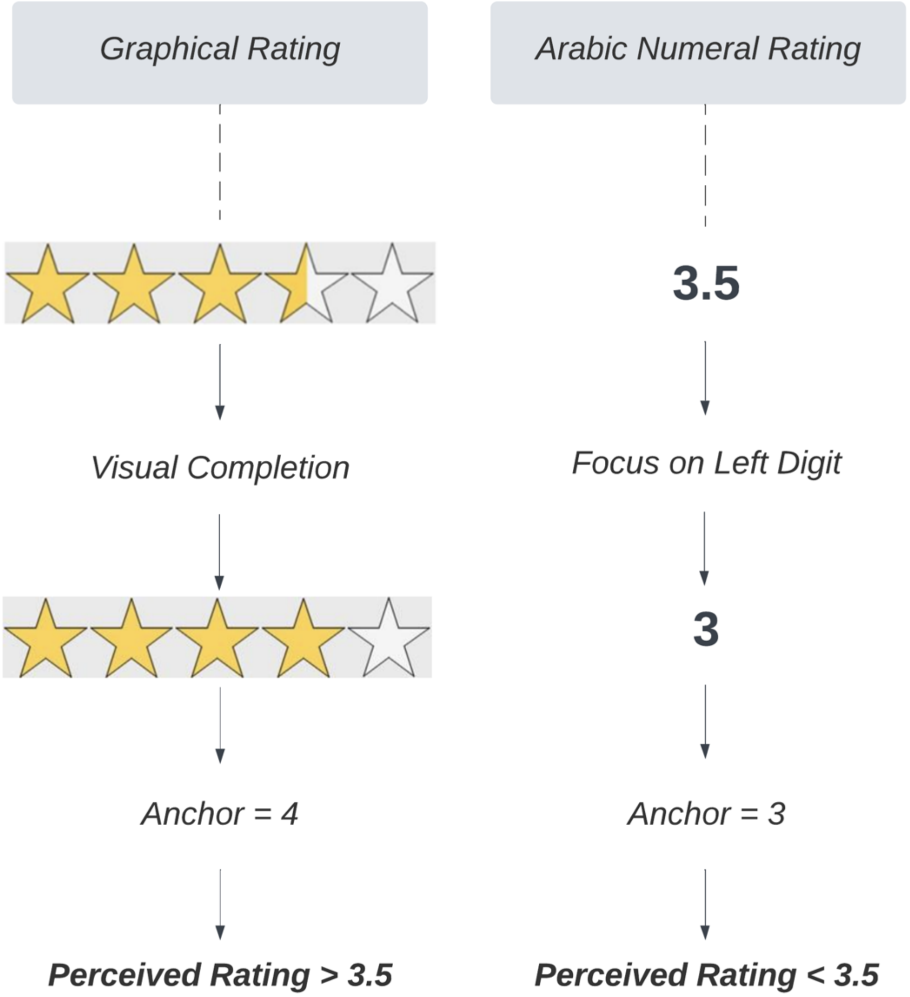
Anchors Differ by Rating Symbols

We posit that perceptual biases can influence the choice of the initial anchor. When presented with graphical symbols, people may be more likely to anchor on rounded-up numbers, leading to an overestimation of the magnitude of fractional ratings. However, when presented with Arabic numerals, people may be more 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 perceptual biases.

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 (Coren, Porac, and Theodor 1986; Kanizsa 1979; van Lier and Gerbino 2015; Pessoa and De Weerd 2003; Pessoa, Thompson, and Noë 1998; Pinna 2008; Zemel et al. 2002). 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).

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 for a schematic depiction of the postulated mechanism.

Figure 1: Initial Anchors Elicited by Fractional Ratings



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 leftmost 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 number in the case of fractional numerical ratings. One of the 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; Macé 2012; Manning and Sprott 2009; Stiving and Winer 1997; Strulov-Shlain 2021). 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

This conceptualization of perceptual biases in ratings yields several novel predictions. First, we hypothesize that star ratings will activate rounded-up numbers as anchors because of visual completion (e.g., 4 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 lower than the fractional number (e.g., 3 will be the anchor for 3.5). Thus,

H1: Ratings presented using graphical symbols will be overestimated, while equivalent ratings presented in Arabic numerals will be underestimated.

Second, our conceptualization of the effect of perceptual biases on perceived magnitude identifies an important boundary condition. We predict that visual completion will increase the perceived magnitude of fractional star ratings, but not whole numbered star ratings, since there is no scope for visual completion for the latter (e.g., four full stars). Similarly, we expect that left-digit bias will reduce the perceived magnitude of fractional Arabic numerals, but not whole numbered Arabic numerals, as the left digits of the latter are not different from the number (e.g., 4.0). Thus,

H2: Overestimation of graphical ratings and underestimation of Arabic numeral ratings will manifest for fractional ratings, but not for whole ratings.

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. In fact, since marketing platforms display ratings that are averaged over several round ratings, we conjecture that fractional ratings will be more prevalent than round ratings.

Finally, our conceptualization also suggests a way to mitigate the bias caused by visual completion of star ratings. The root cause of the overestimation of graphical ratings is the automatic propensity to complete visually incomplete pictures. Thus, if we use visually complete pictures (see Figure 2) to depict the fractional ratings, the perceptual system will not try to complete the graphical symbols.

H3: The overestimation of graphical ratings will be alleviated if the graphical symbols used to describe them are visually complete.

Alternate Accounts & Empirical Package

Apart from the visual completion of incomplete graphical symbols, there are other mechanisms that could lead to the overestimation of fractional star ratings. We have identified two plausible alternative mechanisms. First, people may simply round the fractional ratings to the nearest whole number due to cognitive shortcuts or miserliness. Such rounding of fractional ratings could lead to overestimation of the perceived magnitudes of ratings with fractions greater than half (e.g., overestimation of 3.75), but it cannot account for the overestimation of fractions lower than half (e.g., overestimation of 3.25). Second, consumers may overestimate the magnitude of fractional star ratings because they associate star shapes with positive connotations, and this positive association is mistakenly attributed to magnitude judgments. This explanation can account for the overestimation of star ratings but not for overestimation of graphic ratings that use circles. We investigate these alternative accounts in our empirical studies.

We conducted six laboratory studies (N = 2,612) to test our predictions. All the studies were preregistered 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). Studies 1 and 2 tested the tendency to overestimate numerical quantities represented using graphical symbols and to underestimate them when represented using Arabic numerals. Studies 3 and 4 delved deeper into the underlying mechanisms of these perceptual biases. Study 5 explored whether the perceptual biases in numeric ratings influence consumers' memory recalls. Finally, Study 6 explored the practical implications of these patterns in magnitude judgments by examining consumers' willingness to pay and quality judgments of products presented with graphical or Arabic numeral ratings.

Figure 2: Rating Symbols Used in Studies 1-6

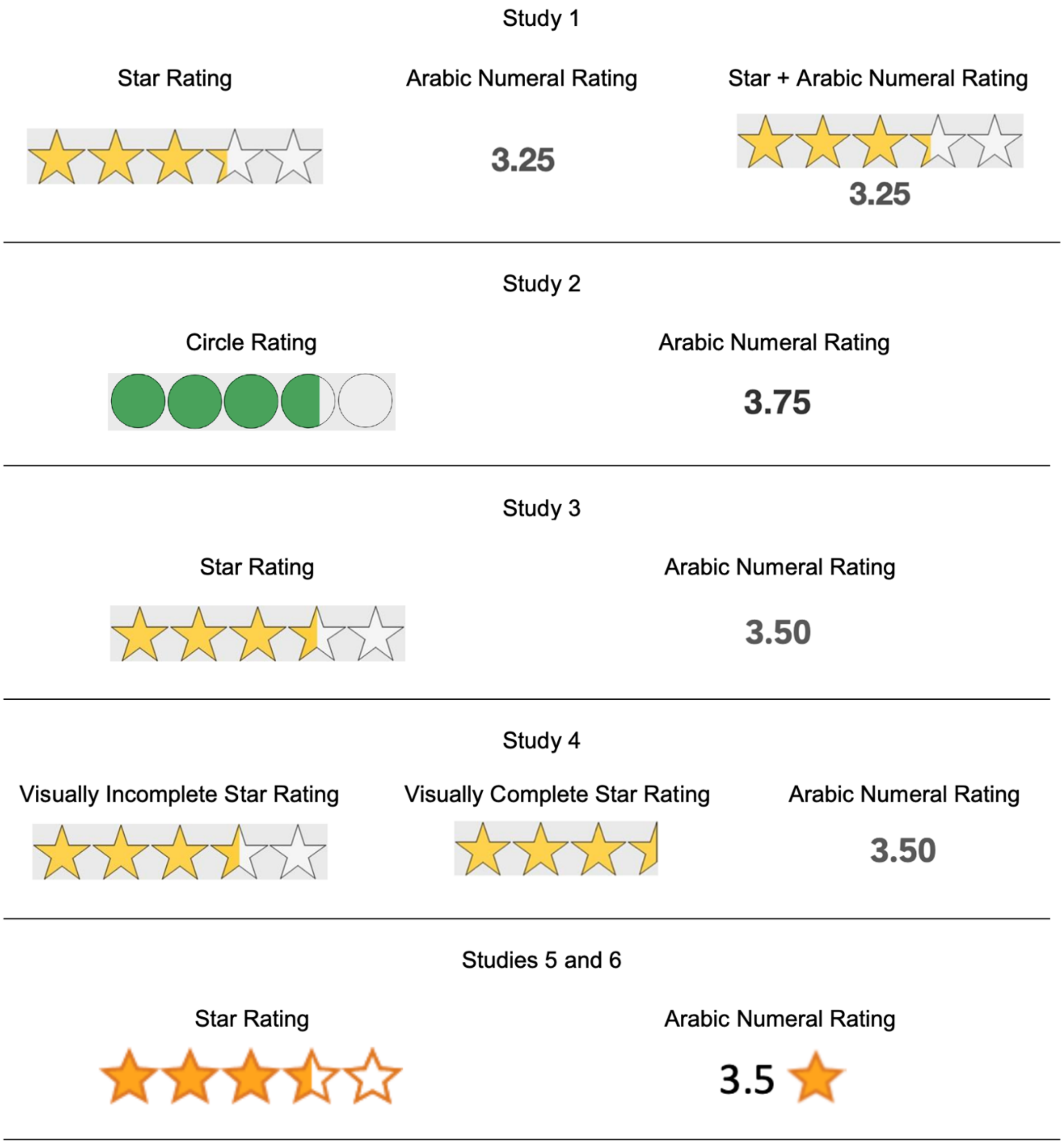


Figure 2—This figure shows the symbols used for numeric ratings in Studies 1-6. In graphic ratings, the magnitude of the rating is represented by the percentage of the area that is colored. For example, a rating of 3.25 would be represented by three fully colored shapes and a fourth shape that is only 25% colored.

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. Specifically, we aimed to test H1 and H2 along with exploring the effect of multiple representations (i.e., star and Arabic numerals together) on the subjective magnitude perception.

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 (Barth and Paladino 2011; Booth and Siegler 2008; Siegler and Opfer 2003; Siegler and Ramani 2009; Slusser, Santiago, and Barth 2013). We adapted this task to test our hypotheses. Participants were shown several numbers and asked to indicate their perceived magnitude on a number line. The numbers were presented either as star symbols, Arabic numeral symbols, or both star and Arabic numeral symbols (see Figure 2). This study was pre-registered at OSF (http://bit.ly/3Wtri2k).

Participants and Procedure

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_{age} = 41 years).

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 Web Appendix for details). The numeric ratings in calibration phase and in 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 (Figure 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 Web Appendix for a visual illustration). The ratings included five whole numbers (1.00, 2.00, 3.00, 4.00, 5.00), four fractions with quarters (1.25, 2.25, 3.25, 4.25), four fractions with halves (1.50, 2.50, 3.50, 4.50), and four fractions 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 anchored at 1 on the left and 5 on the right.

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 graphical depiction of means of perceived magnitudes across the three conditions. Figure 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 star ratings and underestimated the magnitude in the 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, fractional ratings were more likely to be biased than whole number ratings.[1]

[1] It might be argued that the lack of bias for whole numbers is because we used these numbers—1, 3, and 5—in the calibration phase. However, we do not find any bias 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 round numbers.

Figure 3A: Perceived Magnitude of Ratings (Study 1)

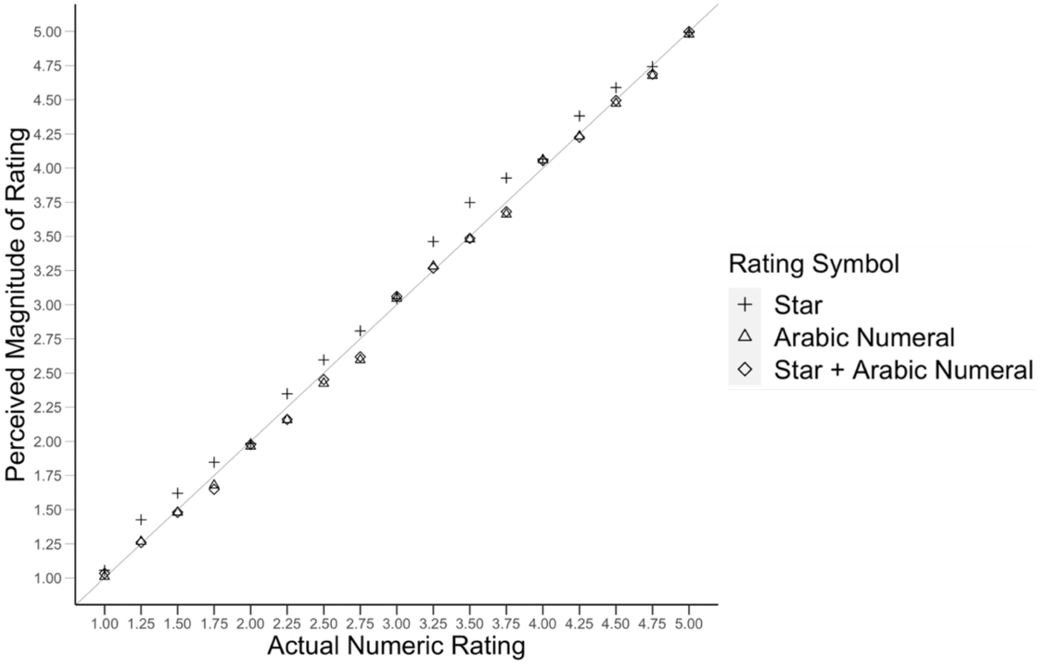


Figure 3B: Bias in Magnitude Perceptions (Study 1)

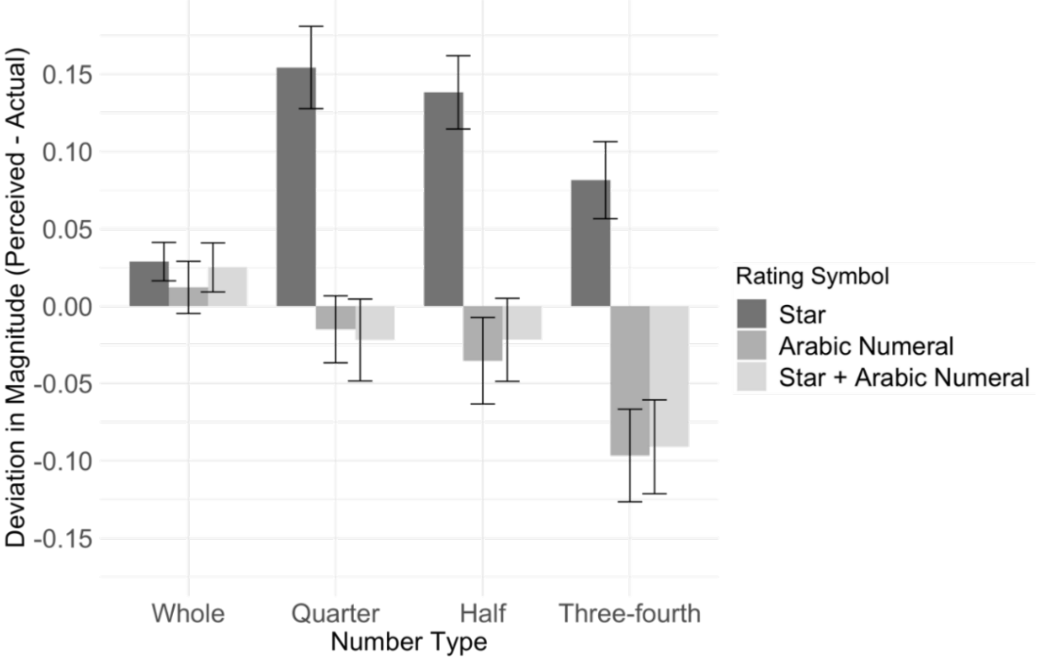


Figure 3A and 3B—These figures show the results from Study 1. Figure 3A shows how the average perceived magnitudes of the 17 stimuli change with different rating symbols. Figure 3B shows the bias in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) by type of number across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

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 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.), fractions with quarters (1.25, 2.25, etc.), fractions with halves (1.50, 2.50, etc.), and fractions 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 (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 factor[2]. The significant main effect of rating symbol ($F(2, 613) = 52.76$, $p < .0001$, $\eta^2_p = .15$) was qualified by a significant two-way interaction ($F(6, 1839) = 38.15$, $p < .0001$, $\eta^2_p = .11$). Figure 3B displays the deviation of subjective magnitude perception from the actual magnitude of the ratings for these number types.

Participants overestimated star ratings ($M_{\text{Star}} = +.10$, 95% CI = [+.08, +.12]), but underestimated Arabic numeral ratings ($M_{\text{ArabicNumeral}} = -.03$, 95% CI = [-.05, -.01]) and combined ratings ($M_{\text{Star+ArabicNumeral}} = -.03$, 95% CI = [-.05, -.01]). There was no significant difference in participants' responses in the Arabic numeral and the combined rating conditions ($t(613) = -.43$, $p = 0.67$) across all number types.

The bias in the perception of star ratings was higher for incomplete star ratings ($M_{\text{FractionalStars}} = +.12$, 95% CI = [+.10, +.15]) than for whole star ratings ($M_{\text{WholeStars}} = +.03$, 95% CI = [+.00, +.05]; $t(1839) = 10.92$, $p < .0001$). This suggests that only visually incomplete stars are overestimated, while visually complete stars are less likely to be overestimated. Similarly, the

[2] 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.

bias in the perception of Arabic numeral ratings was greater for fractional numeric ratings; fractional Arabic numerals were underestimated ($M_{\text{FractionalArabicNumeral}}$ = -.05, 95% CI = [-.07, -.03]), while whole numerals were less likely to be underestimated ($M_{\text{WholeArabicNumeral}}$ = +.01, 95% CI = [-.01, +.04]; $t(1839) = -6.99$, $p < .0001$).

We also investigated how the bias in perception of fractional ratings changed with distance from the anchors. For fractional star ratings, the bias in magnitude perception was higher for ratings with quarter stars ($M_{\text{.25Stars}}$ = +.15, 95% CI = [+.13, +.18]) than for ratings with three-fourths stars ($M_{\text{.75Stars}}$ = +.08, 95% CI = [+.06, +.11]; $t(1839) = 6.78$, $p < .0001$). This suggests that the bias increases when the actual numbers are farther from the rounded-up numbers. Fractional star ratings are instinctively rounded up due to automatic visual completion, and then adjusted downward to correct for the rounding. Because of insufficient adjustment, the bias is larger for fractional stars that are farther away from the initial anchor. Therefore, the bias caused by rounding up and insufficient downward adjustment is strongest for stars with fractional quarters and weakest for stars with fractional three-fourths.

The pattern reversed for Arabic numeral ratings; the bias was higher for three-fourths ratings ending in .75 ($M_{\text{.75ArabicNumerals}}$ = -.10, 95% CI = [-.12, -.07]) than for quarter ratings ending in .25 ($M_{\text{.25ArabicNumerals}}$ = -.02, 95% CI = [-.04, +.01]; $t(1839) = -7.63$, $p < .0001$). This suggests that Arabic numeral ratings are instinctively anchored on the left-most digits (Thomas and Morwitz 2005), and then adjusted upward to correct for the fractions. Because the adjustment is insufficient, the bias is larger for fractional numbers that are farther away from the initial anchor. Therefore, the bias caused by rounding down and insufficient upward adjustment is strongest for fractional three-fourths and weakest for fractional quarters.

Discussion

Several important results emerge from Study 1. First, the study reveals that individuals tend to overestimate the magnitude of numerical quantities represented by star symbols and underestimate those described by Arabic numerals. Thus, H1 is supported. Next, supporting H2, this was found to be true only for fractional numbers and not for whole numbers.

The pattern of results also offers insights into the mechanisms underlying such biases. We found that the overestimation bias in star symbols was stronger for quarter ratings than for three-fourths ratings. Conversely, the underestimation bias 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 is the bias. Thus, these results support our contingent anchoring hypothesis.

Furthermore, Study 1 found that participants' magnitude judgments of ratings using a combination of star and Arabic numerals were found to be similar to those using only Arabic numerals, indicating that Arabic numerals have an advantage during encoding process. That is, when both star ratings and Arabic numerals are available, people tend to pay more attention to the latter, possibly because Arabic numerals are more commonly used to represent numerical quantities and are thus easier for the brain to process. Therefore, from a consumer welfare perspective, these findings suggest that Arabic numeral ratings might be better than star ratings.

But star ratings might be more aligned with marketers' and retailers' goals of showing the products in the best possible light.

Importantly, this study also rules out the alternative hypothesis that participants rounded fractional star ratings to 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.

Study 1 used stars to compare graphical symbols with Arabic numeral symbols. Although such a design has high external validity as many platforms use star ratings, it raises concerns about the generalizability of the observed overestimation bias. Will this effect manifest for graphical ratings that do not use stars? Perhaps stars have a positive halo, which might be biasing people's magnitude judgments. To address this concern, the next study will use a different type of graphical symbol, circles.

STUDY 2: CIRCLES VERSUS ARABIC NUMERALS

Study 2 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 the previous study 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[3]. 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 (45% non-male, M_{age} = 38 years). In Study 2, we employed the same procedure as in Study 1, but with two modifications. First, we substituted stars with circles (see Figure 2), 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.

[3] 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).

Analyses and Results

The findings of this study largely align with those of the previous study (See Figures 4A and 4B). 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 the previous study 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 factor[4]. 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 ($F(3, 1323) = 24.27$, $p < .0001$, $\eta^2_p = .05$).

[4] 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 4A: Perceived Magnitude of Ratings (Study 2)

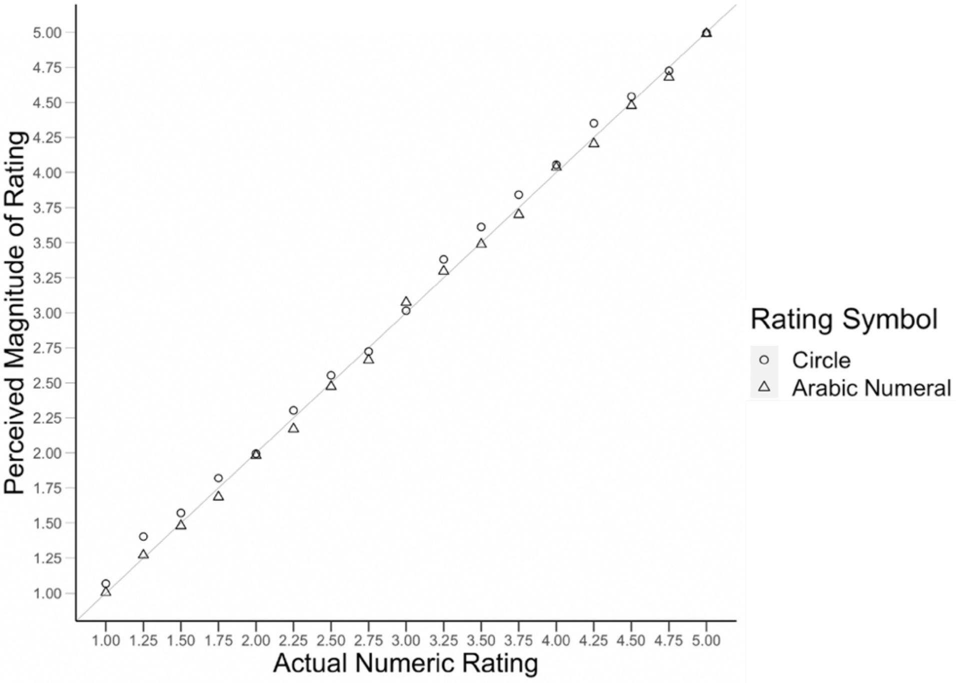


Figure 4B: Bias in Magnitude Perceptions (Study 2)

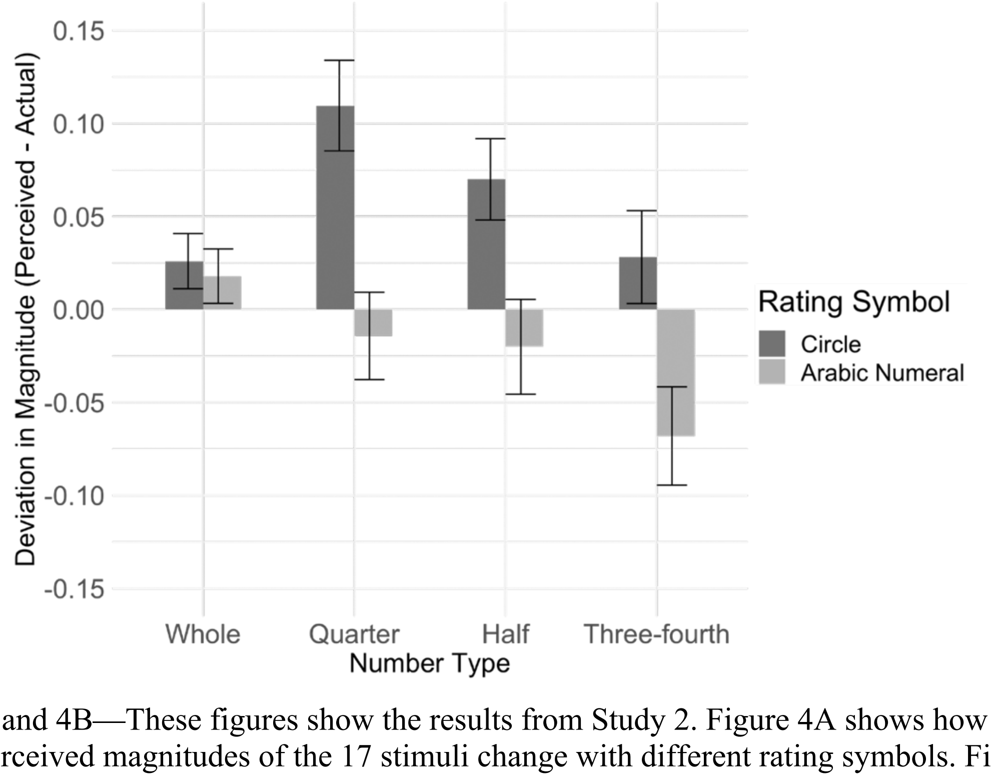


Figure 4A and 4B—These figures show the results from Study 2. Figure 4A shows how the average perceived magnitudes of the 17 stimuli change with different rating symbols. Figure 4B shows the bias in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) by type of number across rating symbols. The bars in Figure 4B represent 95% confidence intervals. Both figures depict raw means.

First, participants overestimated circle ratings (M_{Circle} = +.06, 95% CI = [+.04, +.08]), but underestimated Arabic numeral ratings ($M_{\text{ArabicNumeral}}$ = -.02, 95% CI = [-.04, -.00]; $t(441) = 5.89$, $p < .0001$). Second, the overestimation bias in the perception of circle ratings was more pronounced for incomplete circles ($M_{\text{FractionalCircles}}$ = +.07, 95% CI = [+.05, +.09]) than for complete circles ($M_{\text{WholeCircles}}$ = +.03, 95% CI = [+.00, +.05]; $t(1323) = 5.25$, $p < .0001$). Similarly, the underestimation bias in the perception of Arabic numeral ratings was greater for fractional numeric ratings ($M_{\text{FractionalArabicNumerals}}$ = -.03, 95% CI = [-.05, -.01]) than for whole number ratings ($M_{\text{WholeArabicNumerals}}$ = +.02, 95% CI = [-.00, +.04]; $t(1323) = -6.27$, $p < .0001$).

Third, the bias in the perception of fractional circle ratings increased when the anchor was farther from the actual number. The bias was higher for quarter ratings ending in .25 ($M_{\text{.25Circle}}$= +.11, 95% CI = [+.09, +.13]) than for three-fourths ratings ending in .75 ($M_{\text{.75Circle}}$ = +.03, 95% CI = [+.01, +.05]; $t(1323) = 8.08$, $p < .0001$). This shows that the bias increases when the anchors are farther from the actual ratings. The pattern reversed for Arabic numeral ratings; the bias was more pronounced for three-fourths ratings ending in .75 ($M_{\text{.75ArabicNumerals}}$ = -.07, 95% CI = [-.09, -.05]) than for quarter ratings ending in .25 ($M_{\text{.25ArabicNumerals}}$ = -.01, 95% CI = [-.04, +.01]; $t(1323) = -5.30$, $p < .0001$). This suggests that Arabic numeral ratings are instinctively anchored on the left-most digits, and then adjusted upward to correct for the fractions. Thus, these results once again demonstrate that the bias increases when the anchors are more distant from the actual ratings.

Discussion

Study 2 demonstrates the robustness of the results observed in the previous study (as shown in Figures 4A and 4B). The results indicate that the observed overestimation bias in magnitude judgments of numerical quantities represented using graphical symbols extends to various shapes, including stars and circles.

One limitation of the previous studies is that we do not have direct evidence that people use rounded-up numbers as anchors for star ratings and rounded-down numbers as anchors for Arabic numerals. The next study was designed to explicitly test this assumption in our theory.

STUDY 3: DIFFERENT ANCHORS

In Study 3, we aimed to verify the hypothesis that people tend to instinctively round up fractional star ratings and round down fractional Arabic numeral ratings, creating anchor points that bias 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

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_{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.

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 Web 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.

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 (β = 2.21, p < .0001, d = 1.22). Participants rounded up 79% of the half star ratings (M_{Star} = .79, 95% CI = [.72, .85]). However, only 29% of the same half ratings were rounded up when expressed in Arabic numerals ($M_{\text{ArabicNumeral}}$ = .29, 95% CI = [.22, .37]). The two-way interaction (β = .17, p = .25, d = .09) was not significant, indicating that the effect was consistent across all rating values (see Table 1).

Table 1: % Choice of Rounded-up (vs. Rounded-down) Anchor in Study 3

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—This table shows summary of results for Study 3. The values are estimated from models reported in the text.

Discussion

The results of Study 3 validate our assumption that incomplete stars are instinctively rounded up. Participants tended to round up fractional graphic ratings and round down fractional Arabic numerals. These findings indicate that when presented with fractional graphic ratings, people tend to anchor on the rounded-up value. In contrast, when the same fractional ratings are expressed as Arabic numerals, people tend to anchor on the rounded-down value (i.e., the left-digit of the Arabic numeral rating). The systematic choice of different anchors and insufficient adjustment from these anchors result in the overestimation and underestimation biases observed in magnitude judgments for graphical ratings and Arabic numeral ratings, respectively.

STUDY 4: 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, a perceptual bias, is the cause of the overestimation of fractional star and circle ratings. This bias is activated by visual cues of incompleteness. Previous studies have demonstrated that providing visual cues of completion can alleviate this bias (Gerbino 2020; Kanizsa 1979). 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 does not occur 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 would also reduce the overestimation of fractional star ratings (H3). In Study 4, we aimed to verify the hypothesis.

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 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_{age} = 42 years).

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 (as shown in Figure 2) on an unmarked number line. We used the same paradigm as in the first two studies, 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 biases, 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 to the same extent.

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 and fractional numeric ratings and submitted the average deviations to a 2 x 2 x 2 mixed ANOVA with type of stars (visually complete vs. visually incomplete) as the between-subjects factor and number type (whole vs. fractional) and rating symbol (star vs. Arabic numeral) as within-subjects factors[5].

Two patterns are noteworthy. First, consider the perception of star ratings. The two-way interaction between the type of stars and number type was significant for star ratings ($F(1, 1077.02) = 46.58$, $p < .0001$, $\eta^2_p = .04$). In line with the results of Studies 1 and 2, we found that participants overestimated fractional star ratings in the visually incomplete stars condition ($M_{\text{FractionalStars}} = +.11$, 95% CI = [+.10, +.13]). However, when the stars were visually complete, they did not overestimate the fractional star ratings ($M_{\text{FractionalStars}} = +.02$, 95% CI = [-.00, +.03]. The average overestimation for fractional numbers in the complete star condition was

[5] The three-way interaction was significant ($F(1, 541) = 28.86$, $p < .0001$, $\eta^2_p = .05$).

significantly lower than that in the incomplete star condition ($t(1259) = 8.13$, $p < .0001$); see Figure 5 for a visual representation. These results suggest that the overestimation of fractional star ratings is caused by visual completion. Additionally, the perception of whole star ratings was not affected by the visual manipulation of completeness ($M_{\text{VisuallyIncomplete}} = +.02$, 95% CI = [+.01, +.04]; $M_{\text{VisuallyComplete}} = +.00$, 95% CI = [-.01, +.02]; $t(1259) = 1.83$, $p = .07$).

Figure 5: Visual Completeness Mitigates Overestimation Bias in Fractional Star Ratings (Study 4)

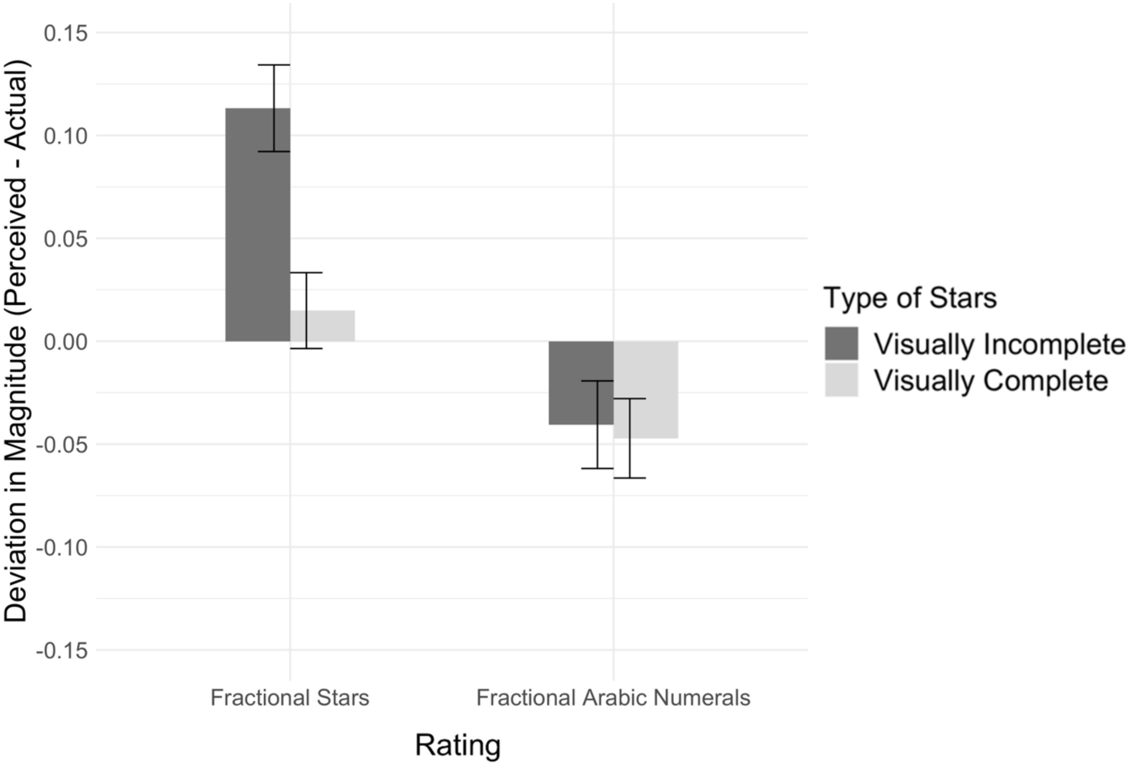


Figure 5—This figure shows the results of Study 4. The plot shows the bias in the subjective magnitude perception (i.e., perceived magnitude – actual magnitude) of fractional star ratings and fractional Arabic numeral ratings by type of stars used in Study 4. The bars represent 95% confidence intervals. The plot depicts raw means.

Now consider the perception of Arabic numerals. There was no effect of the type of stars on the underestimation of Arabic numerals. We found that fractional Arabic numeral ratings were underestimated ($M_{\text{FractionalArabicNumerals}} = -.04$, 95% CI = [-.06, -.02]) when the stars were visually incomplete as well as when they were visually complete ($M_{\text{FractionalArabicNumerals}}$ = -.05, 95% CI = [-.06, -.03]). The extent of underestimation of fractional Arabic numerals did not depend on the type of stars used in the preceding task (*t*(1259) = .55*, p* = .58; see Figure 5). Whole Arabic numerals were not underestimated regardless of the type of star used in the preceding task ($M_{\text{VisuallyIncomplete}}$ = +.03, 95% CI = [+.01, +.04]; $M_{\text{VisuallyComplete}}$ = +.01, 95% CI = [-.00, +.03]; *t*(1259) = 1.02*, p* = .31). These findings indicate that when star and Arabic numeral ratings are presented in sequence, fractional Arabic numeral ratings are still underestimated, regardless of whether the preceding star ratings were overestimated or not.

Discussion

The findings of Study 4 suggest that the overestimation of fractional star ratings is caused by visual completion. Consistent with Studies 1 and 2, we found that participants overestimated fractional star ratings when visually incomplete stars were used. However, when the stars were visually complete, participants did not overestimate the fractional star ratings. Thus, H3 is supported. Study 4 also demonstrates that even after exposure to star ratings, the tendency to underestimate fractional Arabic numeral ratings remains a robust and persistent bias.

The first four studies demonstrate the presence of two perceptual biases in the subjective magnitude perceptions of numeric ratings: the visual completion bias and the left digit bias. However, we haven't investigated if these biases could also impact consumers' recall of the actual value of the numeric rating. We designed Study 5 to shed light on this.

STUDY 5: MEMORY RECALL OF RATINGS

Do the divergent effect of visual completion and left-digit bias carry over to recall? The present study was designed to address this question. In this study, participants saw products with various ratings and were subsequently asked to recall the ratings. Half of them saw star ratings and the other half saw Arabic numeral ratings. We tested whether the type of rating symbol influenced the accuracy of the recall.

Note that people naturally recall ratings in Arabic numerals. This necessarily renders star ratings more susceptible to a bias, and Arabic numerals less so. For Arabic numerals, recall does not entail any transcoding; participants have to only retrieve the encoded digits from working memory. As a result, we did not expect any underestimation biases in Arabic numeral ratings in this study. However, for star ratings, participants had to first convert the star images to Arabic numerals and then retrieve it. This transcoding process makes star ratings more susceptible to anchoring effects caused by visual completion. Thus, we predicted that recall values of star ratings would be overestimated, whereas there would be no systematic underestimation for Arabic numeral ratings. Study 5 was pre-registered at OSF (http://bit.ly/3V7KXnz).

Participants and Procedure

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 conditions[6]: 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_{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.[7] We used 1404 recalled rating responses for the analysis.

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 (Figure 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 Web 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.

[6] 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.

[7] 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.

Analyses and Results

Recall Magnitudes. We conducted a 2 x 4 mixed 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. Our results showed that participants recalled higher rating values when they saw star ratings (M_{Star} = 3.14, 95% CI = [3.10, 3.17]) compared to Arabic numeral ratings ($M_{\text{ArabicNumeral}}$ = 3.01, 95% CI = [2.98, 3.05]). The main effect of the rating symbol was significant ($F(1, 347) = 20.40$, $p < .0001$, $\eta^2_p = .06$), but the interaction was not significant ($F(3, 1044) = .79$, $p = .50$, $\eta^2_p = .00$). See Table 2 for descriptive statistics.

Table 2: Average Recalled Ratings in Study 5

Numeric Rating	**Star Symbols**	**Arabic Numeral Symbols**
1.5	1.70	1.58
2.5	2.69	2.52
3.5	3.62	3.52
4.5	4.52	4.44
Average (3.0)	**3.14**	**3.01**
95% CI	[3.10, 3.17]	[2.98, 3.05]

Table 2—This table shows summary of results for Study 5. The values are estimated from models reported in the text.

Deviation in Recall Magnitudes. To analyze the biases 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 - Actual Magnitude). We then subjected the deviations to a 2x4 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. The analysis results showed a significant overestimation bias in participants' recalls of star ratings, with a mean of +.14 (95% CI = [+.10,

+.18]). However, there was no bias in the recalls of Arabic numeral ratings ($M_{\text{ArabicNumeral}}$ = +.02, 95% CI = [-.02, +.06]).

Discussion

The results of Study 5 suggest that the bias in star ratings extends to memory recalls, while the bias in Arabic numerals does not. Specifically, consumers are more likely to recall inflated ratings when presented with star ratings, rather than Arabic numerals. These findings have important implications for the way consumers remember and share product ratings through word-of-mouth. When consumers rely on their memory to make purchasing decisions or convey their product impressions to others, the use of star ratings by retailers is more likely to result in biased decisions and impressions than the use of Arabic numerals. In contrast, Arabic numeral ratings are less susceptible to such biases, and may be more accurate in memory recall. Thus, when it comes to recalls, Arabic numeral ratings seem more veridical than star ratings.

The first five studies were designed to characterize the mental processes contributing to biases in magnitude judgments of numerical quantities. While the first five studies have provided insights into these biases, the downstream consequences of these biases on consumers' judgments and decisions have yet to be explored. Study 6 was designed to address this gap.

STUDY 6: WILLINGNESS TO PAY

In Study 6, 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 6 was pre-registered at OSF (http://bit.ly/3BJdc5a).

Participants and Procedure

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.

We randomly assigned participants to either the star rating or Arabic numeral rating condition (Figure 2). Participants were shown several ballpoint pens listed on an e-commerce website, along with their average ratings out of five (see Web 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.

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_{Star} = \$2.63, $M_{\text{ArabicNumeral}}$ = \$2.35; $F(1, 325) = 6.85$, $p = .009$, $\eta^2_p = .02$). Additionally, participants rated pens with star ratings to be of higher quality than those with Arabic numeral ratings ($M_{\text{Star}} = 2.48$, $M_{\text{ArabicNumeral}} = 2.38$; $F(1, 332) = 7.60$, $p = .006$, $\eta^2_p = .02$). See Table 3 for details.

Table 3: Willingness to Pay and Quality Judgments in Study 6

	Willingness to Pay ($)		Quality Judgments	
Numeric Rating	**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 3—This table shows summary of results for Study 6. The values are estimated from models reported in the text.

Discussion

Study 6 showed that the type of symbols used to present ratings can influence consumers' judgments and decisions. The findings suggest that consumers may be more inclined to pay higher prices and view products as having superior quality when the ratings for those products are displayed using graphical symbols, rather than Arabic numerals. These results indicate that the perceptual biases in the way people encode numerical ratings have real-life implications and practical significance.

GENERAL DISCUSSION

The use of numerical ratings for evaluation has become widespread in many platforms, such as social media, navigation, and product review sites. It has been shown that numerical ratings can have a significant impact on sales, with even small increases leading to increased product sales (McKinsey 2021). The format of the ratings, however, varies across platforms. Some platforms use graphic symbols, like stars or circles, to represent ratings, while others use Arabic numerals. Despite the wide prevalence of numerical ratings and their influence on consumer behavior, research on the effects of rating formats on consumers' judgments is scarce. The current research is the first to examine the differential effects of graphical and Arabic numeral ratings on consumers' evaluative judgments.

Studies 1 and 2 show that there is a systematic bias in people's perception of numeric ratings. Ratings presented using graphical symbols (e.g., image of three and a half stars) are overestimated, while ratings using Arabic numerals (e.g., 3.5) are underestimated. Furthermore, we found that the overestimation in graphical ratings and underestimation in Arabic numeral ratings manifested only for fractional ratings, and not for whole ratings. Additionally, we observed that overestimation bias in graphical ratings was stronger for quarter ratings (e.g., image of three stars and a quarter star) compared to three-fourth ratings (e.g., image of three stars and a three-fourth star), and in contrast, the underestimation bias in Arabic numeral ratings was stronger for three-fourth ratings than for quarter ratings. These observations support the contingent anchoring account of the biases in magnitude judgments of fractional ratings. The studies show that the perceived magnitudes of fractional ratings are contingent on the anchors activated and the distance between the anchors and the fractional numbers.

Studies 3 and 4 test the assumptions of this account. Study 3 tests this account and shows that rating symbols indeed influence the choice of the anchor used in the encoding of ratings, such that people instinctively round up fractional star ratings and round down fractional Arabic numeral ratings.

We hypothesize that visual completion of graphic symbols contributes to the overestimation of graphical ratings. To test this, Study 4 uses two types of star symbols, visually complete versus incomplete, and finds that fractional star ratings are overestimated when presented with visually incomplete star symbols, but not when presented with visually complete star symbols. Study 5 shows that the visual overestimation bias can even affect consumers' recall of actual numeric ratings. Finally, Study 6 reveals that the impact of ratings symbols on the magnitude judgments of ratings influences consumers' judgments and decision making.

Theoretical Implications

Our research adds to the existing literature on the intersection of visual perception and numerical cognition, making significant contributions in these areas. First, previous research has suggested that humans process and retain visual representations of objects more effectively than verbal representations (Lieberman and Culpepper 1965; Paivio, Rogers, and Smythe 1968). However, our research findings reveal that when it comes to processing numerical quantities, verbal representations such as Arabic numerals may be easier to process than graphical representations such as star symbols. According to the results of Study 1, when participants were presented with numerical ratings in both graphical symbols and Arabic numerals, they tended to rely primarily on the Arabic numerals to estimate the magnitudes. These findings suggest that, in some circumstances, verbal representations such as Arabic numerals may facilitate the brain's processing of stimuli.

Second, previous research on the processing of multi-digit fractional numbers has suggested that people tend to focus heavily on the round number (i.e., the left digit) and ignore the fractional digits (i.e., the digits to the right of the decimal point) (Stiving and Winer 1997). For instance, when assessing the magnitude of a number like 2.75, individuals may primarily attend to the left digit (i.e., 2), and neglect the digits after the decimal point (i.e., 7 and 5). However, our findings contradict this view. We observed that when people encode a multi-digit fractional number, such as 2.75, they use both the left and right digits in forming their subjective perceptions of magnitude. Specifically, they anchor their judgment on the left digit, but also take into account the right digits in shaping their final magnitude estimate, thus indicating that subjective perception of magnitude depends on both the anchor and the distance between the anchor and the actual number. These results extend the existing literature on left-digit bias.

Managerial Implications

Our findings can help managers in their decision-making regarding rating symbols. Utilizing graphical symbols, instead of traditional Arabic numerals, could improve consumer perceptions of product quality and potentially result in higher conversion rates, lower cart abandonment, and heightened product expectations. However, study 1 results show that when star ratings and Arabic numerals are presented side-by-side, people generally attend to Arabic numerals only, suggesting that Arabic numerals may be easier to process. This indicates that the human perceptual system prefers Arabic numerals to stylized graphical representations of quantities. Furthermore, study 5 shows that recalls of Arabic numerals are more veridical. These findings suggest that from a consumer welfare perspective, Arabic numeral ratings might be better than star ratings. But star ratings might be more aligned with marketers' goals of promoting products.

Our research also provides a key input for managers considering the benefits and drawbacks of using graphical symbols relative to Arabic numerals. Our findings showed that the superior evaluation of graphical symbols over Arabic numerals only applies to fractional ratings, not whole ratings. To fully understand the advantage of graphical symbols, managers should take into account the distribution of fractional and whole ratings on their platform, as higher frequency of fractional ratings might suggest more managerial benefits of using graphical rating symbols.

Limitations and Future Research

Our results also identify several promising areas for further exploration. First, the external validity of the effect of choice of rating symbols discussed in this research are yet to be confirmed. Our experiments were conducted in a controlled environment, but online platforms, such as online marketplaces, are complex and dynamic environments with multiple factors at play. Therefore, our findings need to be further validated using field studies or archival data. Such studies will also help us assess the true effect size of our phenomenon, outside laboratory settings.

Second, we used stars and circles in our research to demonstrate the visual completion bias in graphical rating symbols. However, past research has shown that the extent of visual completion can vary across shapes (Jia, Wan, and Zheng 2022). Hence, it is important to understand the variability of overestimation biases across different types of graphical symbols. Future research could compare different shapes to determine which shapes elicit the highest and lowest overestimation biases in magnitude judgments of numerical quantities.

Third, we have not examined the impact of rating symbols on the ease of comparing products. It is possible that it is easier to compare products using Arabic numeral ratings than using graphical ratings. This difference in ease of comparison could potentially affect consumers' product choices. To explore this possibility, conducting conjoint studies that vary the rating symbols and measure their effects on consumer choices would be useful.

In conclusion, our research demonstrates that the type of symbolic representation used has an impact on the magnitude judgment of numeric ratings, and that this is influenced by different types of perceptual biases. These results underline the need to further explore perceptual processes in numerical cognition.

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.

Barth, Hilary C. and Annie M. Paladino (2011), “The Development of Numerical Estimation: Evidence Against a Representational Shift,” *Developmental Science*, 14 (1), 125–35.

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.

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.

Coren, Stanley, Clare Porac, and Leonard H. Theodor (1986), “The Effects of Perceptual Set on the Shape and Apparent Depth of Subjective Contours,” *Perception & psychophysics*, 39 (5), 327–33.

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.

Epley, Nicholas and Thomas Gilovich (2004), “Are Adjustments Insufficient?,” *Personality and Social Psychology Bulletin*, 30 (4), 447–60.

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.

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 (2022), “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.

Macé, Sandrine (2012), "The Impact and Determinants of Nine-Ending Pricing in Grocery Retailing," *Journal of Retailing*, 88 (1), 115–30.

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 15, 2023), [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.

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.

Pinna, Baingio (2008), "A New Perceptual Problem: The Amodal Completion of Color," *Visual neuroscience*, 25 (3), 415–22.

PowerReviews (2021), "Survey: The Ever-Growing Power of Reviews," *PowerReviews*, (accessed December 7, 2022), [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.

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.

Sevilla, Julio, Mathew S. Isaac, and Rajesh Bagchi (2018), "Format Neglect: How the Use of Numerical Versus Percentage Rank Claims Influences Consumer Judgments," *Journal of Marketing*, 82 (6), 150–64.

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.

Siegler, Robert S. and Geetha B. Ramani (2009), "Playing Linear Number Board Games—But Not Circular Ones—Improves Low-Income Preschoolers' Numerical Understanding," *Journal of Educational Psychology*, 101, 545–60.

Slusser, Emily B., Rachel T. Santiago, and Hilary C. Barth (2013), "Developmental Change in Numerical Estimation," *Journal of Experimental Psychology: General*, 142, 193–208.

Stiving, Mark and Russell S. Winer (1997), "An Empirical Analysis of Price Endings with Scanner Data," *Journal of Consumer Research*, 24 (1), 57–67.

Strulov-Shlain, Avner (2021), "More than a Penny's Worth: Left-Digit Bias and Firm Pricing," *Chicago Booth Research Paper*, (19–22).

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.

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.

Zemel, Richard S., Marlene Behrmann, Michael C. Mozer, and Daphne Bavelier (2002), "Experience-Dependent Perceptual Grouping and Object-Based Attention.," *Journal of Experimental Psychology: Human Perception and Performance*, 28 (1), 202.

WEB APPENDIX

Contents

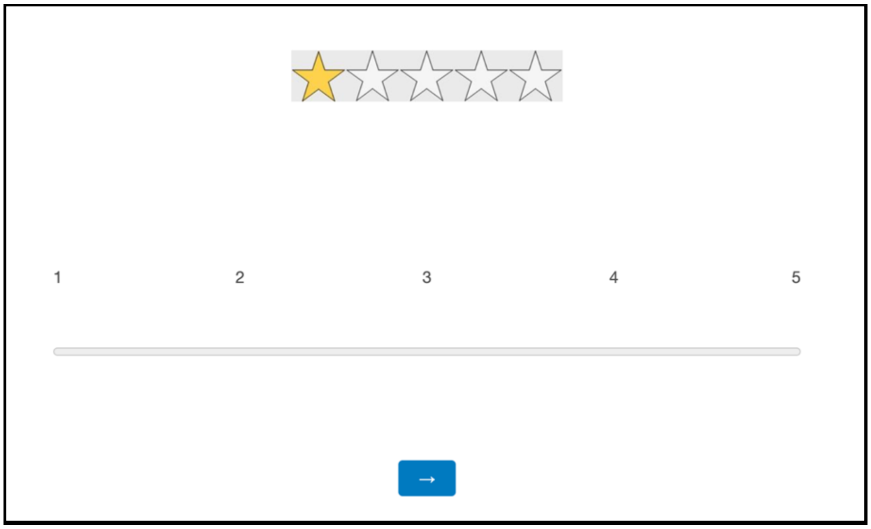
Disclosure: These materials have been supplied by the authors to aid in the understanding of their paper.

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.

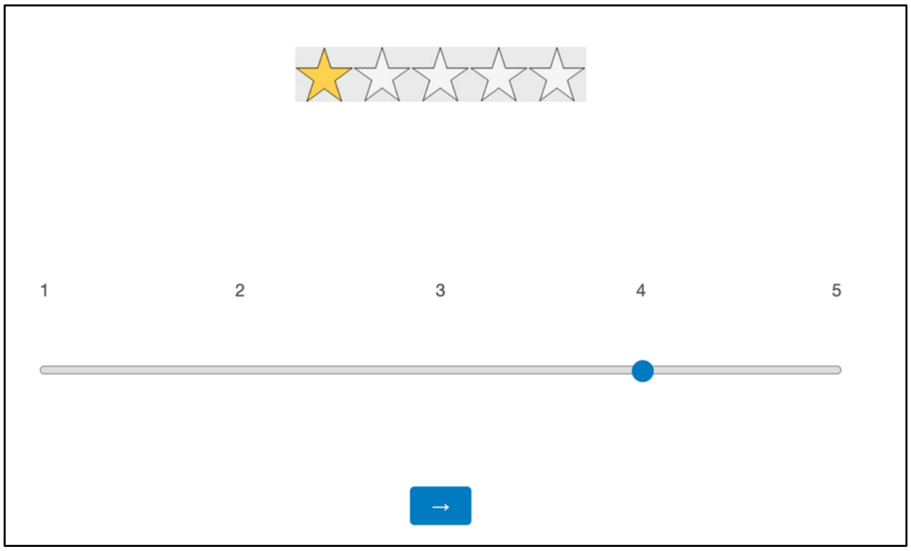
STUDY 1 STIMULI

1. *Calibration Phase*– Star Rating Condition

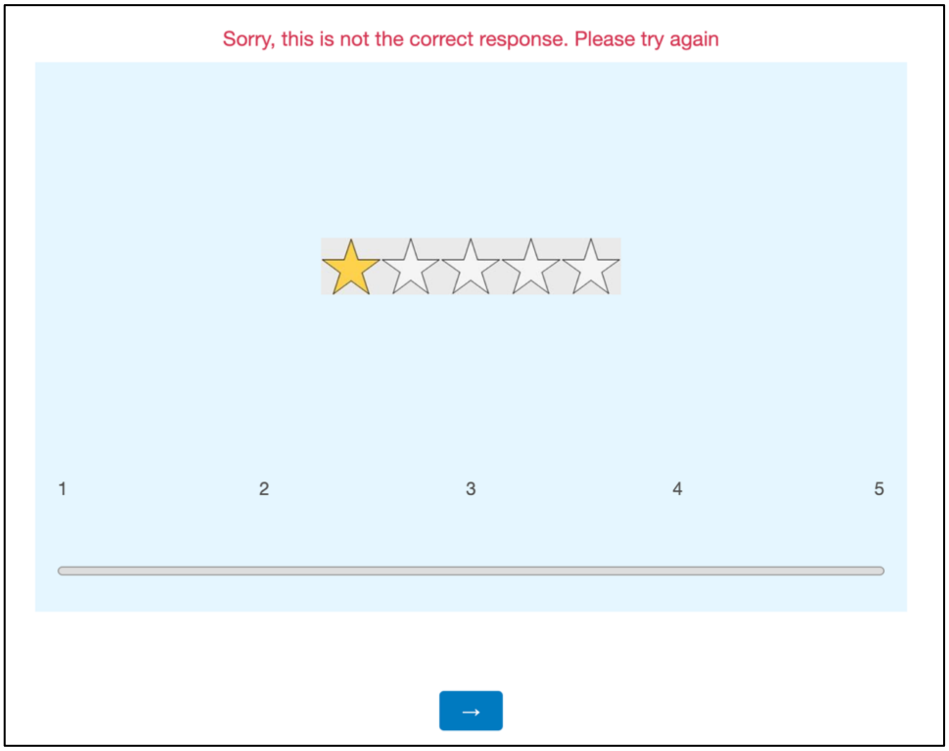
   Participants were administered three questions in the calibration phase. Following is the first question. It was followed by two more questions for values of 3 and 5.



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.



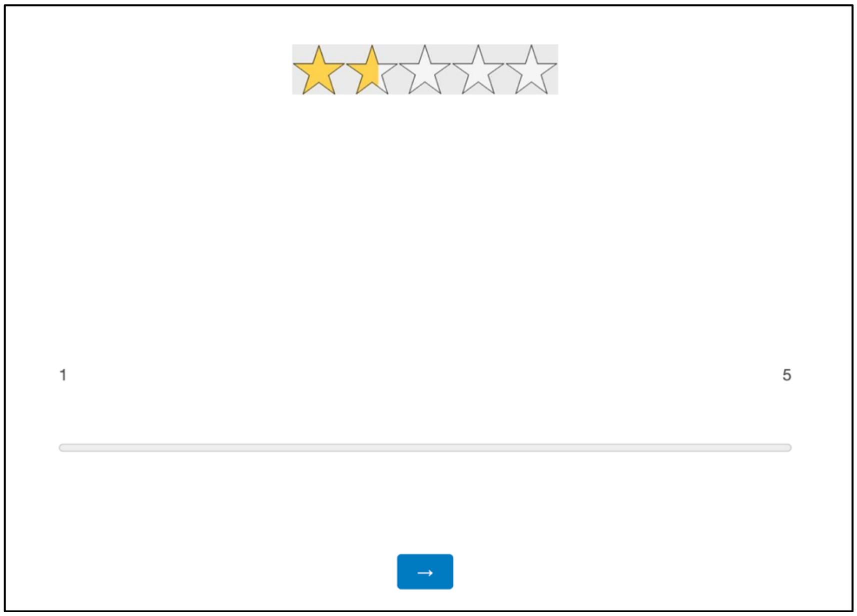
An error message appeared if participant submitted an incorrect response, and the participant was asked to give their response again until they responded correctly.


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

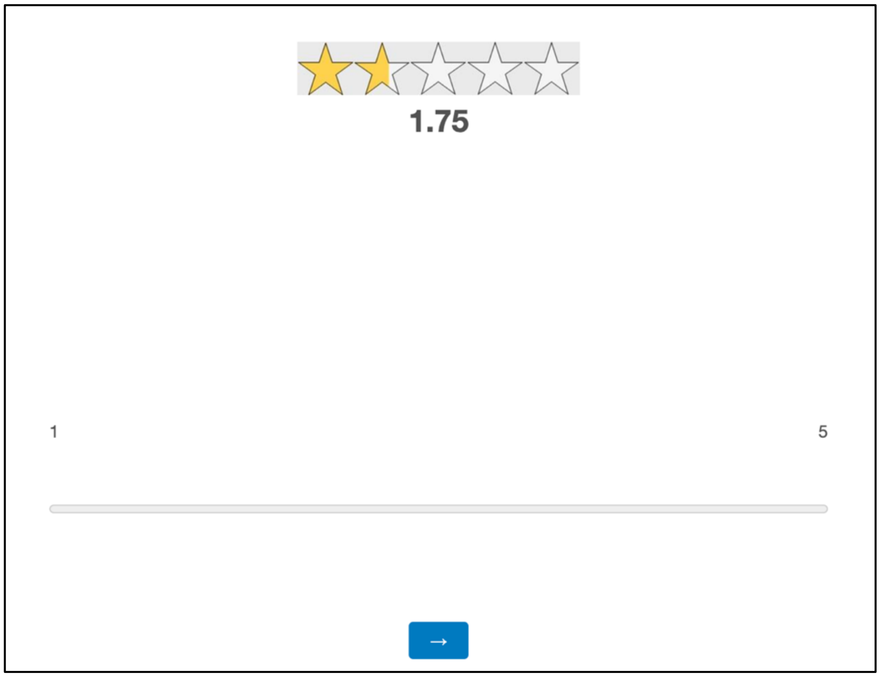
2. *Estimation Task*

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.

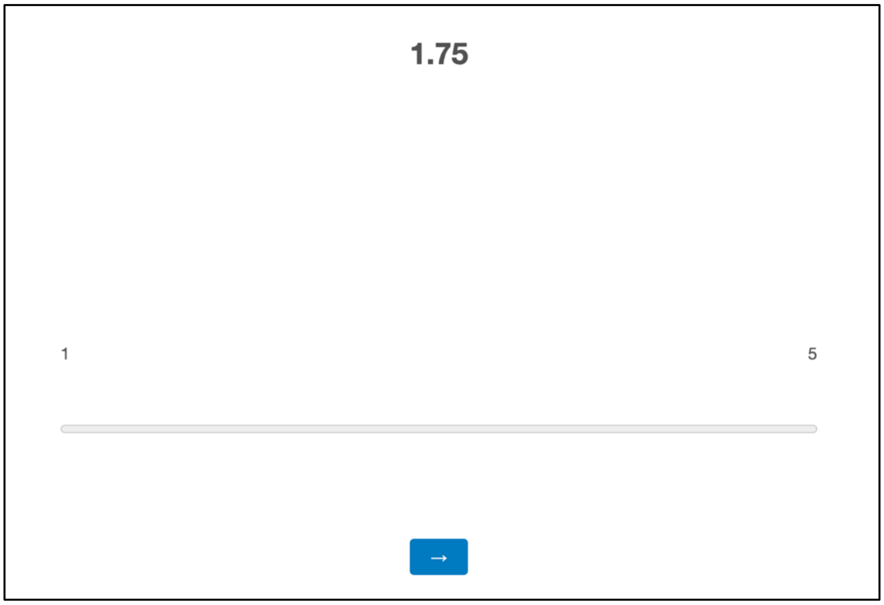
Star Rating Condition (rating = 1.75)



Star + Arabic Numeral Condition (rating =1.75)

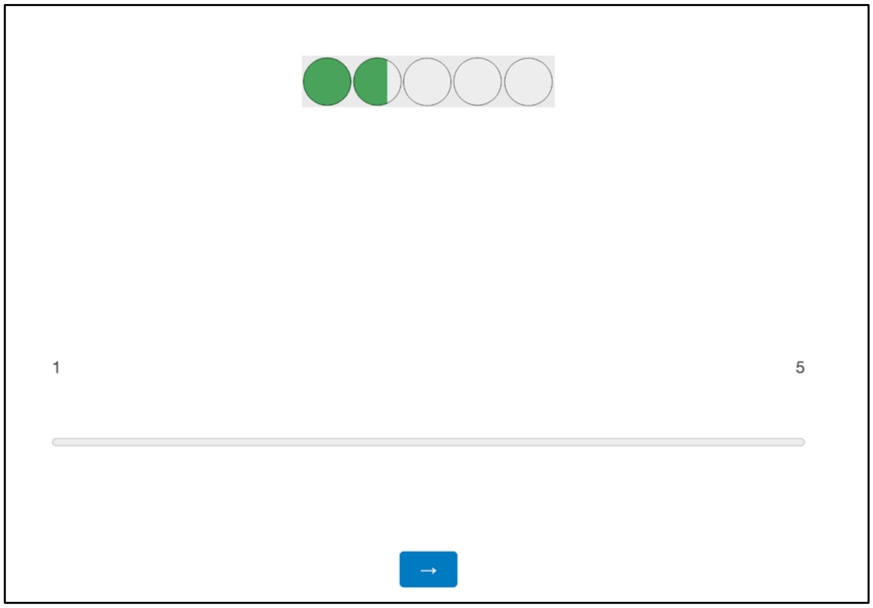


Arabic Numeral Condition (rating =1.75)



STUDY 2 STIMULI

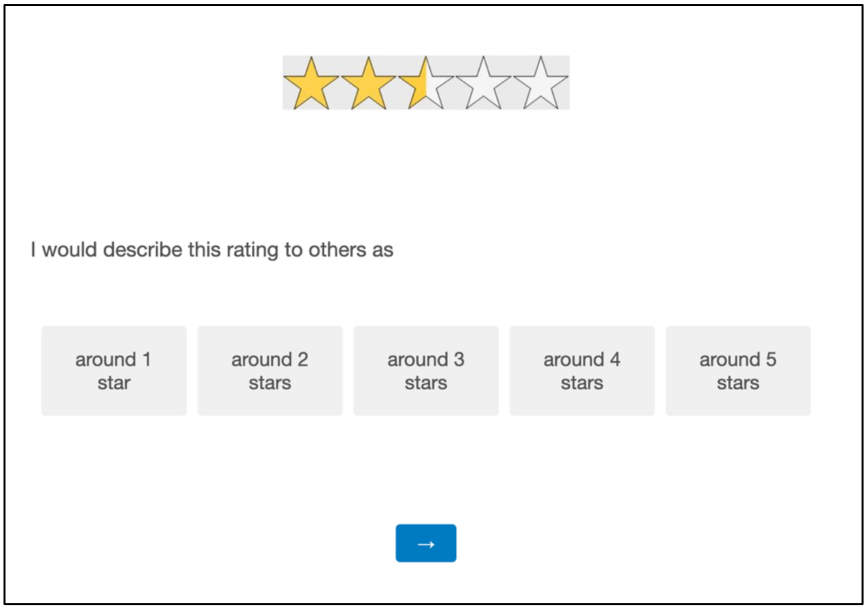
The instructions and paradigm of Study 2 was the same as Study 1. Following is one of the examples of the 17 questions administered in the estimation task in the circle condition.



STUDY 3 STIMULI

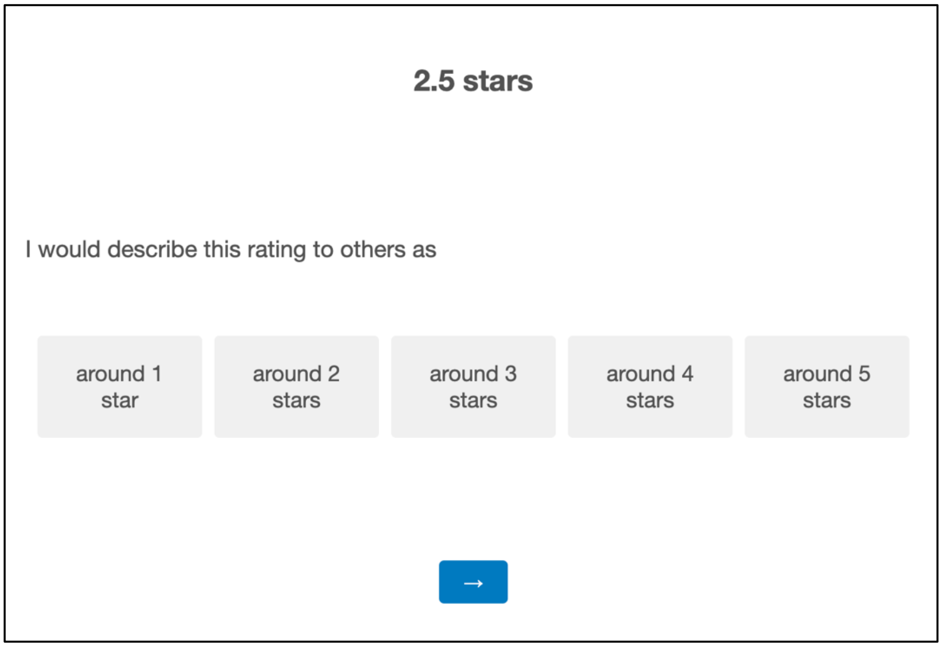
Participants answered four questions in each condition. The order of the four questions was randomized. Each question was displayed on a separate screen.

Star Rating Condition (rating = 2.5)



I would describe this rating to others as

Arabic Numeral Condition (rating = 2.5)



2.5 stars

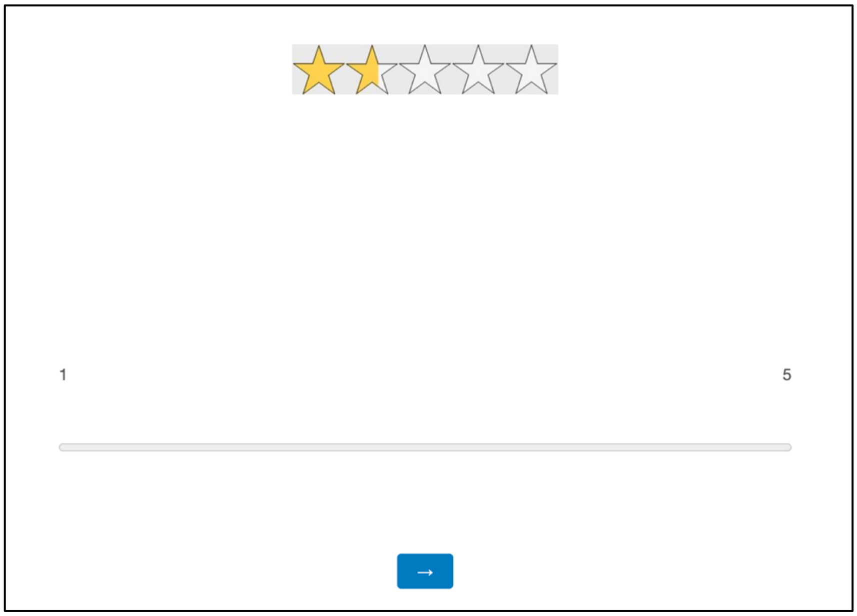
I would describe this rating to others as

STUDY 4 STIMULI

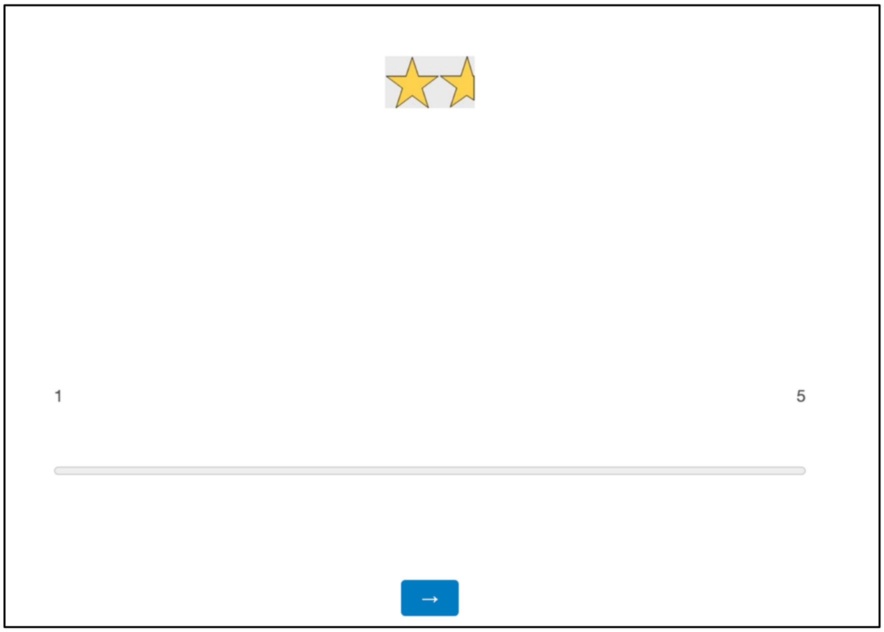
The instructions and paradigm of Study 4 was the same as Study 1.

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)



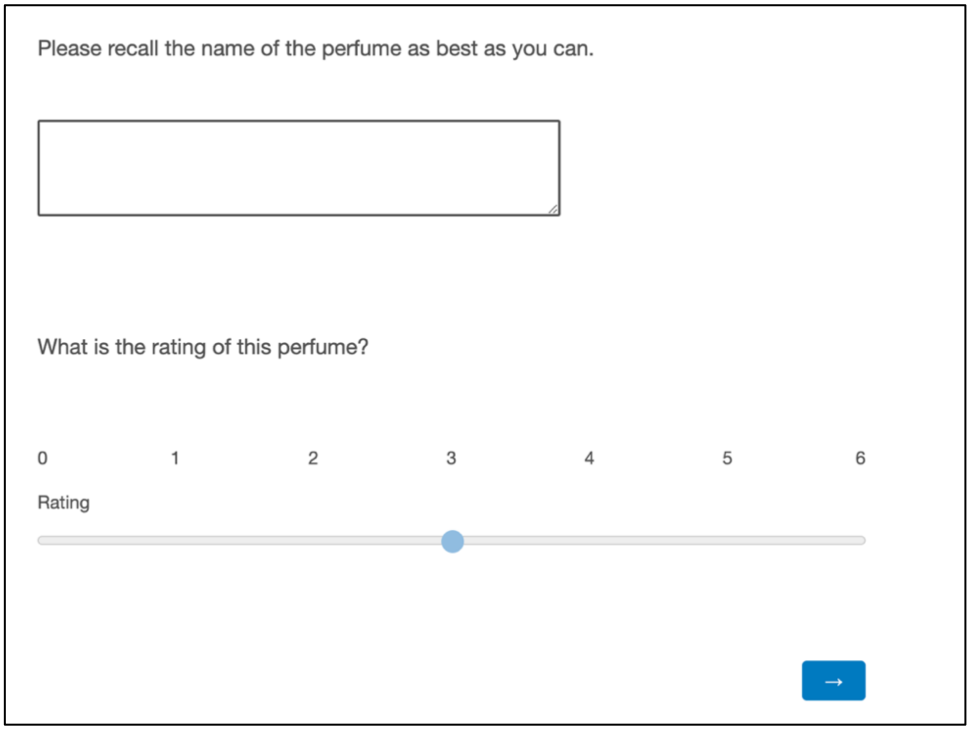
Visually complete star condition (rating = 1.75)



STUDY 5 STIMULI

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.

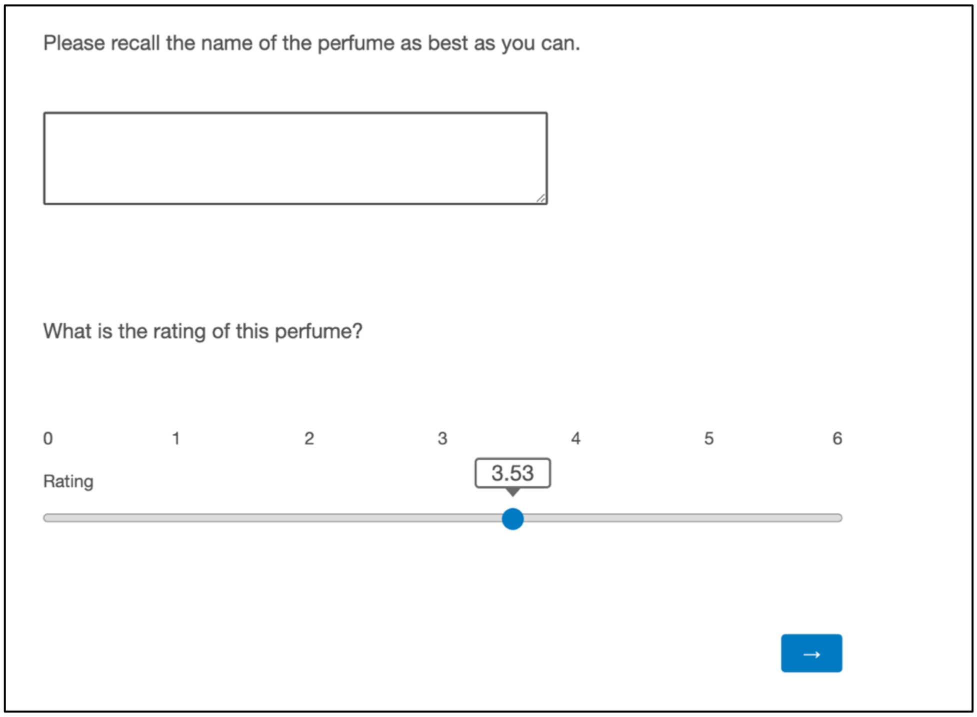
Second screen.



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

What is the rating of this perfume?

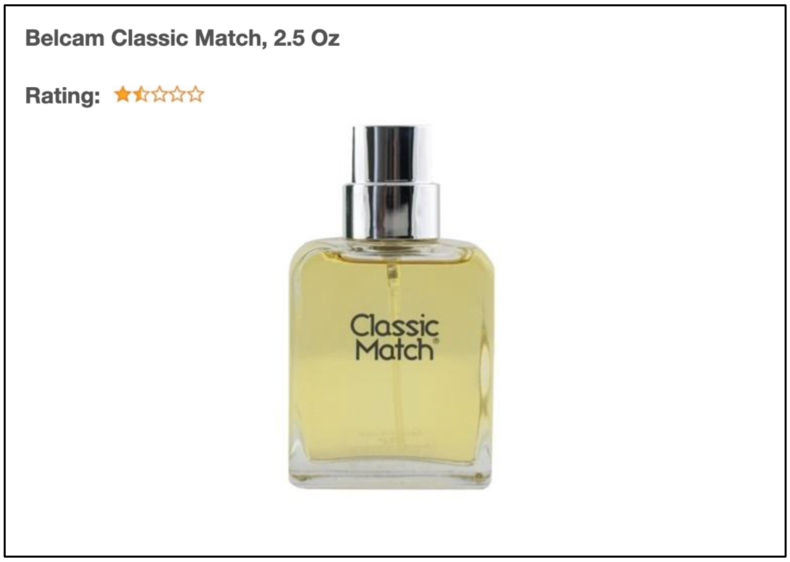
Participants could move the cursor by dragging it. They could see the numeric value of the position of the cursor as they dragged the cursor.



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

What is the rating of this perfume?

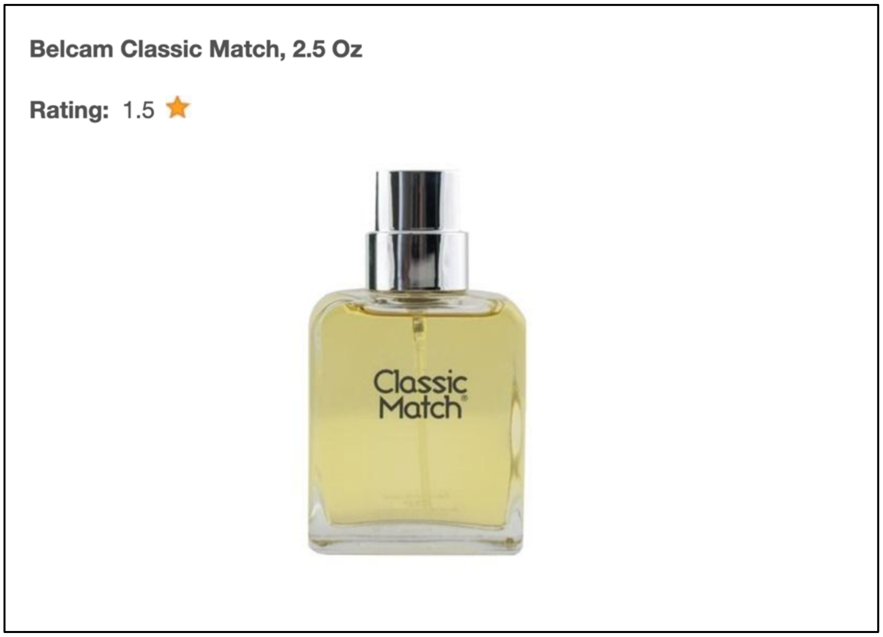
Star Rating Condition (rating =1.5), first screen



Belcam Classic Match, 2.5 Oz

Rating: ★⯪☆☆☆☆

Arabic Numeral Condition (rating =1.5), first screen



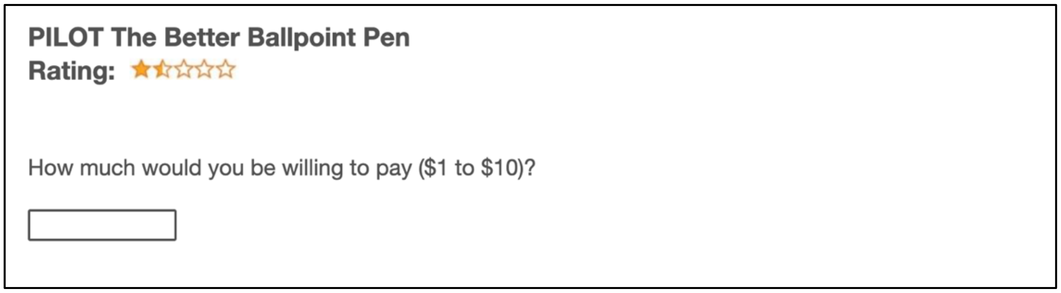
Belcam Classic Match, 2.5 Oz

Rating: 1.5 ★

STUDY 6 STIMULI

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

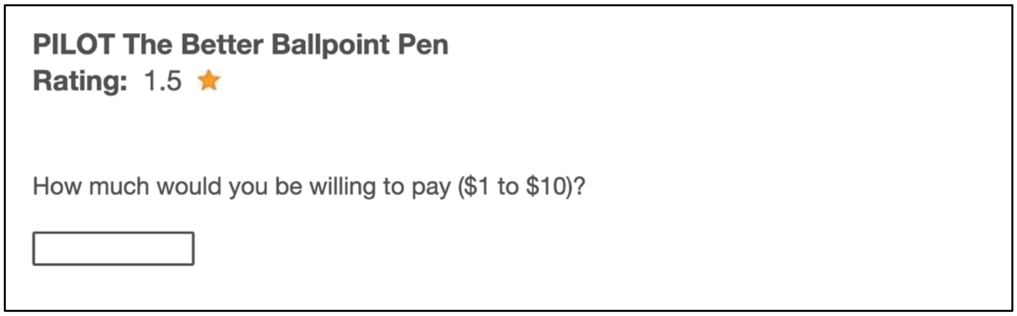
Star Rating Condition (rating = 1.5)



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

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

Arabic Numeral Condition (rating = 1.5)

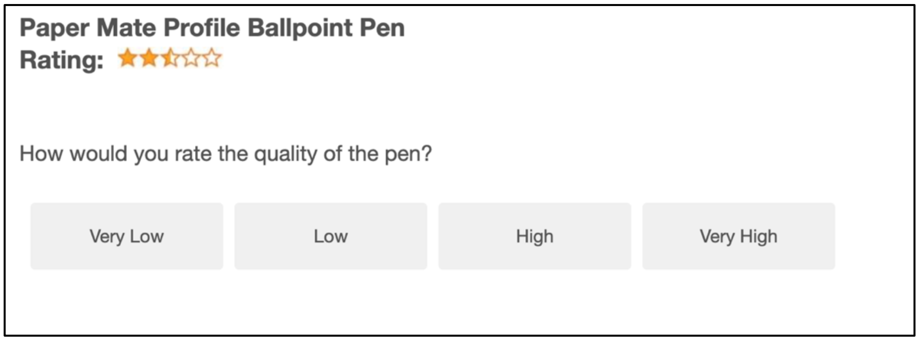


PILOT The Better Ballpoint Pen
Rating: 1.5 ★

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.

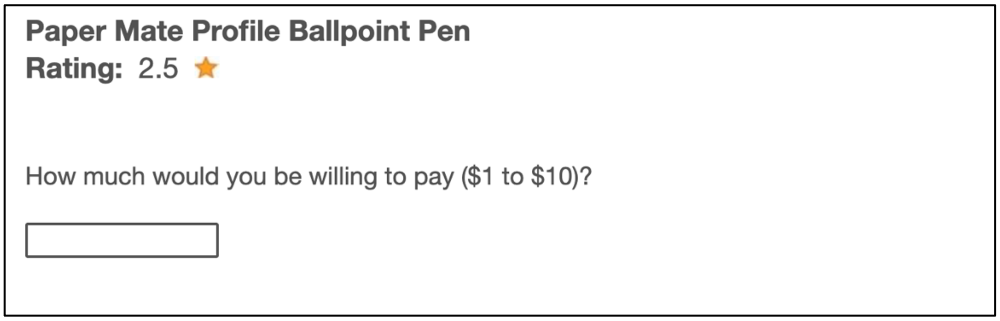
Star Rating Condition (rating = 2.5)



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

How would you rate the quality of the pen?

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)?