Subtraction or Division: Evaluability Moderates Reliance on Absolute Differences versus Relative Differences in Numerical Comparisons

DENGFENG YAN

Specifications of many product attributes (prices, review scores, fuel efficiency, calories, computer processor speed, etc.) are numerical. When comparing alternatives, consumers often need to judge how much larger or smaller one value is than another (say x and y). How do they make such a judgment? The literature suggests that people can rely on either the absolute difference ($x - y$) or relative difference (x / y). Importantly, relying on the absolute versus relative difference might lead to divergent outcomes. Therefore, this research aims to identify one factor that moderates consumers' reliance on absolute versus relative differences. Specifically, we propose that when the attribute is easy to evaluate (i.e., when consumers have clear reference information), people tend to compute and rely on absolute differences to make comparative judgments. By contrast, when the attribute evaluability is low (i.e., reference information is lacking), they tend to compute the relative difference. Results from six studies provide convergent evidence for this proposition and demonstrate its downstream effects on preference and judgments.

Keywords: numerical comparison, absolute versus relative differences, evaluability

Many product attributes are expressed quantitatively. Examples include prices, calories, review scores, mileage per gallon, product warranties, battery life, sun protection factor (SPF), annual percentage rate, hotel room size, and computer processor speed. Therefore, to form preferences and choices, consumers often need to compare two or more numerical values—that is, to judge the degree to which one value is higher than another. For example, how much better is a restaurant with a rating of 4.7 than another with a rating of 4.1? How much more effective is a sunscreen of SPF 30 than another of SPF 15? How deep is a discount from $339 to $299? How much unhealthier is 30 grams of sugar than 20 grams of sugar?

This research explores the psychological process underlying numerical comparisons as in the above examples. Intriguingly, the literature has demonstrated two ways by which consumers form such judgments. The first way is to subtract one number from another—that is, to compute the absolute difference (Biswas et al. 2013; Monga and Bagchi 2012; Wertenbroch, Soman, and Chattopadhyay 2007). The second way is to divide one number by another—that is, to compute the relative difference (Hsee et al. 2009; Palmeira 2011). Then, a question naturally arises concerning when individuals will rely on the absolute or relative difference. Understanding this question is important for theoretical and practical reasons. Theoretically, investigating this question has the potential to shed new light on how

consumers process numerical information, an area that has received increasing attention in the consumer literature (Adaval 2013). Practically, relying on the absolute versus relative difference in numerical comparisons might lead to divergent outcomes such as preference reversals. To provide an example, Pandelaere, Briers, and Lembregts (2011) demonstrated that the perceived superiority of a 9-year product warranty to a 7-year warranty looms larger when the same difference is framed as 108 versus 84 months. This finding is based on the assumption that people compute the absolute difference. If people rely on the relative difference, framing by year or month should not make a difference ($9 / 7 = 108 / 84$).

The chief value and focus of the current research therefore lies in identifying one possible boundary condition under which people rely on the absolute versus relative difference in numerical comparisons. Building on prior work on attribute evaluability (Hsee 1996, 2000), we propose that when the attribute is easy to evaluate (i.e., when consumers have clear reference information), people tend to compute and rely on the absolute difference to make comparative judgments. In contrast, when the evaluability is low (i.e., reference information is lacking), they tend to compute the relative difference (i.e., ratio) and use 1 as a reference point to form their comparative judgment. The results from six studies provide convergent evidence for this proposition and demonstrate its downstream effects on preference and judgments.

This research advances knowledge in two areas: numerical cognition and evaluability. Specifically, with respect to the numerical cognition literature, the current work ties together two streams of research to propose a unified model that elucidates how individuals assess the difference between two numerical values. Relatedly, our findings also contribute to each line of literature by identifying and demonstrating evaluability as a moderator. In addition, the present work not only draws on research on evaluability (Hsee 1996; Hsee et al. 1999) but also adds to that literature in two ways. First, while prior research has shown that evaluability can lead to preference reversal when the decision mode changes from separate to joint, we show that even if the evaluation mode does not shift, attribute evaluability can nevertheless systematically influence consumer preferences through a very different mechanism. Second, our hypotheses and findings suggest that under circumstances in which product attributes are difficult to evaluate, consumers actively make them more interpretable by calculating the ratio between them.

THEORETICAL BACKGROUND

Consider two numerical attribute values: x and y. Consumers often need to judge not only which value is better but also more precisely how much better x is than y. In other words, they need to map the numerical difference onto a psychological scale anchored at "slightly better" to "much better." Existing literature suggests that there are two ways in which consumers can make this judgment. The first way involves computing the absolute difference (i.e., $x - y$), and the second way involves computing the relative difference (i.e., x / y). Below, we first review each set of findings and then introduce our theorizing by elaborating how evaluability can shift consumers toward one type of calculation over another.

Reliance on the Absolute Difference

Several articles suggest that consumers rely on the absolute difference when comparing two options. For example, Wertenbroch et al. (2007) investigated how consumers' perceptions of product value are affected by the face value of money (currency). Participants in one study were asked to imagine making grocery purchases in Spain, and the prices were framed in either euros or Spanish pesetas (166.30 pesetas = 1 euro then). For each product category, participants chose between a store brand and a well-known brand, the latter of which was 15% more expensive. For example, if the store brand was priced at 10 euros or 1,663 pesetas, then the brand version was priced at 11.5 euros or 1,912 pesetas. Their results showed that the likelihood of expensive products being chosen was higher when prices were presented in euros than when they were presented in pesetas. This finding implies that a more numerous currency induced the participants to perceive the well-known brands as being more expensive than store brands. A similar effect was observed after the euro was introduced in France, where French francs are more numerous than euros (Gaston-Breton 2006).

As another example, in a study reported by Burson, Larrick, and Lynch (2009), participants were asked to indicate their preference for two movie rental plans. Compared to plan A, plan B had more new movies, but it was also more expensive. In one condition, the difference in the number of new movies was framed on a per-week basis: plan A and plan B had 7 and 9 movies per week, respectively. In another condition, the same difference was framed on a per-year basis: plan A and plan B had 364 and 468 new movies, respectively. The authors found that the participants favored the superior plan for new movies (plan B) more when the number of new movies was expressed in years than when it was expressed in weeks. Similarly, Pandelaere et al. (2011) found that participants rated the difference between 84 and 108 months as being larger than the difference between 7 and 9 years.

Collectively, these findings converge to show that individuals' perceived difference between two options is amplified when the attribute values are expressed in a more numerous unit, scale, or currency. These results thus imply that participants compute and rely on the absolute

difference when forming comparative judgments. In all these studies, if participants had computed and relied on the relative difference, these effects would not have been observed, as the relative difference was kept constant across conditions. For example, in Pandelaere et al. (2011), the ratio between two warranties remained the same in the year and month conditions (9 / 7 = 108 / 84).

Reliance on the Relative Difference

However, another pattern of results has also been found in the literature, where "people largely evaluate the advantage of one option over another in relative terms" (Hsee et al. 2009). For example, participants in one study reported by Hsee et al. (2009) were asked to choose between two sesame oils, with brand A being more concentrated, and thus more aromatic, than brand B. Brand A was also more expensive. The aroma of these two options was presented differently. In one condition, brand A and brand B had an index value of 9 and 2 (large-ratio specification), respectively, while in another condition, these values were 107 and 100 (small-ratio specification). If participants compute the absolute difference, then these two different specifications should not influence their choice because the absolute difference is the same. However, in line with the ratio assumption, the choice share of brand A in the large-ratio condition was significantly higher than that in the small-ratio condition.

In another study reported by Wong and Kwong (2005), participants were asked to choose between two different Hi-Fi systems. System A could hold fewer numbers of CDs, but it had better sound quality than system B. While the number of CDs was kept constant across conditions (system A: 2 vs. system B: 10), the information regarding sound quality was framed in two ways. One condition used a gain frame—that is, the sound delivery rates for these two systems were 99.997% and 99.99%, respectively. The same information was framed in losses in another condition, such that the signal distortion rates were .003% and .01%, respectively. This framing led to a preference reversal. When the sound quality was expressed in larger numbers (99.997% vs. 99.99%), more participants chose system B because they perceived little difference in terms of sound quality. Conversely, when the sound quality was expressed using small numbers (.003% vs. .01%), the difference in sound quality between the two systems appeared to be much larger to respondents, most of whom then preferred system A (see also Kwong and Wong 2006). These findings, as Hsee et al. (2009) noted, imply participants' reliance on the relative difference.

Finally, some findings could be interpreted either way. For example, Krider, Raghubir, and Krishna (2001) asked participants to indicate the maximum price they were willing to pay for three pizzas of different sizes. The size information was explicitly provided but framed in different ways. In one condition, they told participants that the diameters of the pizzas were 8, 11.25, and 13.75 inches. In another condition, they were told that the areas of the pizzas were 50, 100, and 150 square inches. The results showed that although the objective size variation was the same in both conditions, participants were willing to pay more for a large pizza relative to a small pizza if the sizes of the pizzas were described in terms of the area rather than in terms of the diameter. These findings could be interpreted as evidence for reliance on the relative difference (Hsee et al. 2009). For example, when the diameter increased from 8 to 11.25 inches, the area expanded from 50 to 100 square inches. Thus, the ratio increased from about 1.4 to 2. Nonetheless, the same effect might also be attributed to reliance on the absolute difference, which has been amplified from 3.25 (11.25 − 8) to 50 in the previous example.

Moderating Role of Evaluability

Given that each pattern has been robustly shown in the literature, the question then naturally arises as to what determines whether or the extent to which individuals will compute and rely on the absolute or relative difference when making numerical comparisons. In this research, we do not treat the absolute and relative differences as mutually exclusive. Consumers may compute both differences and integrate them into an overall impression (Shafir, Osherson, and Smith 1993; Wright 2001). Thus, the present work is focused on the relative influence of one type of difference versus the other. Many variables might affect such a tendency. For example, the accessibility of different operations might be one of these factors. Consumers who have just performed subtraction (division) might be more likely to compute the absolute (relative) difference than those who have not. The current work explores the role of evaluability (Hsee 1996; Hsee et al. 1999) as a potential moderator. As we will explain later, this moderator has the potential to accommodate existing findings that we have reviewed previously.

Evaluability. Evaluability reflects the extent to which consumers can make sense—that is, to interpret the desirability—of an attribute value. According to Hsee (1996), an attribute is said to be hard (vs. easy) to evaluate if people do not know how good or bad a given value on the attribute is when the value is presented alone, although they often know which direction of the attribute is better. For example, number of entries is an important attribute for dictionaries, and people know that a higher number of entries is more desirable. However, for most consumers, this attribute is difficult to evaluate. For example, how good or bad is 20,000 entries? By contrast, attributes such as the weight and height of a person are more evaluable. Most individuals can easily understand the meaning of someone weighing 100 or 300 pounds.

The evaluability of an attribute is not an endogenous, fixed property; rather, it largely depends on individuals' knowledge. For example, people who are familiar with the GMAT are able to judge whether 650 is a good score, but those who are unfamiliar with this test will find this attribute difficult to assess. In addition, it also depends on whether the attribute value is presented in isolation or in context. For example, if two dictionaries have 10,000 and 20,000 entries, respectively, people readily know that having 20,000 entries is more desirable. Therefore, attribute evaluability is essentially "all about reference information" (Hsee and Zhang 2010), including "which value is the best possible," "which value is the worst possible," and "what the value distribution of the attribute is" (Hsee et al. 1999). Knowledge, presentation format, and other factors are different sources of reference information.

The evaluability of an attribute can lead to preference reversals: people often heavily rely on easy-to-evaluate attributes when the item is judged in isolation, whereas when two options are evaluated side by side simultaneously, the weight attached to difficult-to-evaluate attributes increases. For instance, in one study, participants were asked to evaluate two job candidates for a computer programmer position. Candidate A had a GPA of 4.9 and had written 10 programs, whereas candidate B had a lower GPA of 3.0 but had written 70 programs. In this example, GPA is easier to evaluate than the number of programs written. When each candidate was assessed separately, the participants showed a greater preference for candidate A. However, this preference was reversed in the joint evaluation, where B was judged more favorably (Hsee 1996).

However, although these findings imply that individuals' reliance on difficult-to-evaluate attributes increases when the judgment mode shifts from a separate to a joint evaluation, how such difficult-to-evaluate attribute information is utilized by individuals in joint evaluations (i.e., comparative judgments) remains unclear. For example, in the previous job candidate example, individuals know that candidate B is more productive than candidate A, but they do not know *how much* more productive candidate B is than candidate A. This is precisely the question that this research attempts to address. In addition, we also aim to understand how consumers respond to easy-to-evaluate attribute information in comparative settings. For example, how do they interpret the difference between a GPA of 3.0 and a GPA of 4.9?

Evaluability and Reliance on the Absolute versus Relative Difference. We suggest that the concept of evaluability has direct implications for consumers' reliance on the absolute versus relative difference in numerical comparisons. Intuitively, when evaluability is high, where people can easily interpret the desirability of a single attribute value, they should also be able to make sense of the absolute difference. For example, a person who knows how

good or bad a GMAT score of 650 is should have little trouble understanding how big or small the difference between 650 and 750 is. By the same logic, when evaluability is low, where people have difficulty interpreting the desirability of a single attribute value, they also should not be able to make sense of the absolute difference. Then how do people comprehend the numerical difference under low evaluability? As we discussed previously, compared to high evaluability, low evaluability implies a lack of reference information. Previous work has shown that when an explicit reference point is missing, people may construct their own references—for instance, by relying on nearby round numbers (Pope and Simonsohn 2011) or contextual information (Adaval and Monroe 2002). Much of the literature, however, has examined the reference points that people construct under separate evaluations. Less is known about what kind of references people use in comparative judgments.

We suggest that under such circumstances, people create their own reference point of 1 and make judgments relative to this reference point. In their everyday lives, individuals often make such comparative judgments, and they should have learned the distribution of relative differences. For example, consumers may have learned that, on average, the discount level is about 10–12% (Biswas et al. 2013). Therefore, they are able to judge how attractive a 5% (50%) discount is. To further illustrate our premise, reconsider the job candidate study mentioned earlier. Why would participants favor candidate B (GPA: 3.0, 70 programs) in joint evaluations? We suggest that they might have perceived the productivity of candidate B as seven times higher than that of candidate A (GPA: 4.9, 10 programs). Although one might not know the distribution of programming performance among all job candidates, one does know what a value seven times higher means.

In summary, we predict that attribute evaluability influences the extent to which consumers rely on the absolute versus relative difference in numerical comparisons. Specifically, when evaluability increases, people's reliance on the absolute (relative) difference increases (decreases). Importantly, the moderator that we propose seems to be consistent with prior literature. On the one hand, we find that the attributes used in the absolute difference literature tend to be easy to evaluate. These attributes include the warranty of dishwashers and calories of snacks (Pandelaere et al. 2011), the prices of groceries (Wertenbroch et al. 2007), and the number of new movies per week for movie rental plans (Burson et al. 2009). On the other hand, product attributes used in the relative difference literature are relatively difficult to evaluate. For example, although evaluability was not explicitly embedded in their conceptualization, Hsee et al. (2009) mentioned that their focus was on situations under which consumers "are unfamiliar with the provided specifications and do not know how to translate these numbers to their consumption experience" (953).

The attributes used include the fictitious aromatic index mentioned earlier, the number of diagonal pixels of digital cameras, a fictitious attribute called SVI for cellphones, and the thickness of potato chips. Similarly, the attributes used in Wong and Kwong (2005) and Kwong and Wong (2006)—such as the signal distortion rate of a Hi-Fi system, proficiency in a fictitious computer program called CY, and color display realism in LCD monitors—are relatively more difficult to evaluate. Finally, it must be acknowledged that although the foregoing analysis is in line with our theorizing, such interpretation is not conclusive because none of these studies have directly varied evaluability and have measured reliance on the absolute versus relative difference. Furthermore, the different results might be attributable to other factors associated with different attributes or product categories.

Overview of Studies

This article reports six studies that systematically evaluated our theory and its downstream effects. Studies 1, 2, and 3 examine the foundational proposition that if attribute values are easy (difficult) to evaluate, consumers are likely to compute the absolute (relative) difference. Specifically, study 1 measured evaluability and showed that participants' tendency to use the relative difference decreases as evaluability increases. Studies 2 and 3 sought to test our proposition using an indirect approach. The rationale underlying these two studies is that if participants indeed compute the absolute or relative difference depending on evaluability, then they should solve the same calculation faster (study 2) and recall the difference in the same form more accurately (study 3). By obtaining results in support of these predictions, studies 2 and 3 increase our confidence in our theorizing. Building on such supportive evidence, studies 4–6 examine the downstream effects of our core proposition on consumer choice and judgments. Specifically, studies 4 and 5 show that prior findings assuming participants' reliance on the absolute or relative difference are moderated by evaluability. Study 6 adds to previous studies by extending from choices to continuous judgments and evaluating the underlying mechanism.

It is important to note that in each study, our focus is on how evaluability affects reliance on the absolute or relative difference and the downstream consequences across conditions. In other words, we do not have a priori predictions regarding participants' reliance on the absolute or relative difference within each condition. For example, in the low-evaluability condition, we do not predict that participants' reliance on the relative difference is greater than their reliance on the absolute difference. There are two reasons. First, this is a calibration issue, depending on the strength of the evaluability manipulation. Second, as we mentioned previously, evaluability is only one of the factors affecting

individuals' reliance on the absolute versus relative difference.

STUDY 1

Method

Study 1 sought to initially test our hypothesis by measuring evaluability. Undergraduate participants ($N = 207$, 108 females) were informed that the study concerned price perceptions. We used countertop ovens as the stimulus because they are neither too familiar nor too unfamiliar. Participants were presented with the prices of two countertop ovens (brand A: $100, brand B: $150) and asked to rate how much more expensive brand B is than brand A using a seven-point scale (1 = a little, 7 = a lot). To measure evaluability, we followed Hsee (1996) in asking participants to indicate the extent to which they had any idea about how high or low the prices were by using a seven-point scale ranging from 1 (I have no idea) to 7 (I have a clear idea). Finally, all participants responded to the key dependent measure. Specifically, they were instructed to fill in the blank in the following sentence: "Brand B is more expensive than brand A by _____." Intuitively, they could respond with either "$50" or "50%." Our focus was on whether their tendency of using a dollar amount or percentage would vary as a function of evaluability—that is, the extent to which they can directly make sense of the prices.

Results and Discussion

Participants' responses were coded as either the absolute difference (e.g., $50) or relative difference (e.g., 50%, or half). Four participants provided incorrect answers (e.g., a third, $40). The current analyses included all responses, and excluding these four observations did not change the current conclusion. Overall, 23% (48 / 207) of the responses were in relative difference. Participants' reported evaluability ranged from 1 to 7, with the average being 3.77 (SD = 1.62). To test our hypothesis, we conducted a logistic regression in which participants' likelihood of responding using the relative difference (absolute difference = 0, relative difference = 1) was regressed on self-reported evaluability. This analysis revealed that, in line with our reasoning, the effect is negative and significant ($B = -.361$, SE = .113; Wald = 10.20, $p = .001$), meaning that participants who found the prices to be less (vs. more) evaluable were more likely to respond in relative differences. The result of bootstrapping with 1,000 resampling revealed that this effect is robust, with a 95% confidence interval excluding zero (−.57, −.18).

Study 1 therefore provides preliminary support for our hypothesis. Nevertheless, one limitation of this study is that evaluability was measured rather than manipulated. In addition to addressing this weakness, the subsequent two studies also tested our theorizing from different angles.

STUDY 2

Study 2 sought to test our proposition using another approach while manipulating evaluability. Specifically, if participants indeed tend to compute the absolute difference or ratios depending on evaluability, then they should later solve the same calculation faster because of prior practice.

Method

Design and Participants. Participants ($N = 211$, 112 females) recruited from Amazon Mechanical Turk (MTurk) were randomly assigned to the conditions using a 2 (evaluability: low vs. high) $\times$ 2 (target calculation: subtraction vs. division) design.

Procedure. The study comprised two parts. The first part was described to have the aim of understanding how consumers perceive prices. Given this pretense, all participants were presented with the prices of two yoga mats (mat A: $17, mat B: $51). Participants in the high-evaluability condition were also informed that the average price of yoga mats was about $20, whereas those in the low-evaluability condition did not receive such information. All participants were asked to indicate how much more expensive B is than A using a seven-point scale (1 = a little, 7 = a lot). We did not observe any significant difference between the low- and high-evaluability conditions ($Ms = 6.36$ vs. 6.28; $F < 1, p > .40$).

Participants then moved to an ostensibly unrelated study that aimed to assess mathematical proficiency. Given this pretense, participants were instructed to solve a calculation problem as fast as they can. Participants who submitted the correct answer fastest had a chance to win a $50 gift card. The calculation problem used in the subtraction (division) condition was 51 – 17 (51 / 17).

Results

Twenty-two participants submitted incorrect answers. We included these observations in the current analyses, and excluding them did not change the results. Participants' response latency ranged from 2.94 to 51.18 seconds, with the average being 12.72 (SD = 7.87) seconds.

We noted that participants' response latency was significantly skewed (skewness = 1.84; Shapiro-Wilk statistic = .84, $p < .001$). We therefore conducted a 2 (evaluability: low vs. high) $\times$ 2 (target calculation: subtraction vs. division) ANOVA on the log-transformed response latency to test our predictions. This analysis revealed a significant interaction only ($F(1, 207) = 9.73$, $p = .002$, $d = .43$). We conducted contrast analyses to explore the meaning of the interaction effect. Specifically, if low evaluability indeed induces participants to calculate the ratio, then participants in the low-evaluability condition

should solve the division problem more quickly than those in the high-evaluability condition. In support of this hypothesis, the response latency to solve the division problem was shorter in the low-evaluability condition (raw: $M = 10.90$, SD = 4.87; log-transformed: $M = 2.28$, SD = .50) than in the high-evaluability condition (raw: $M = 14.87$, SD = 10.18; log-transformed: $M = 2.50$, SD = .62; $t(207) = 2.13$, $p = .035$, $d = .39$). By contrast, participants in the low-evaluability condition (raw: $M = 14.74$, SD = 9.22; log-transformed: $M = 2.53$, SD = .55) spent more time solving the subtraction problem than their counterparts in the high-evaluability condition (raw: $M = 11.36$, SD = 6.59; log-transformed: $M = 2.29$, SD = .52; $t(207) = 2.28$, $p = .02$, $d = .45$).

STUDY 3

Study 3 further tested our premise from a memory perspective. Specifically, we asked participants to recall the absolute difference or relative difference (%) after a delay. Their recall accuracy should be enhanced if the question type matches the actual calculation that they had performed. That is, participants should be more likely to accurately recall the absolute difference when evaluability is high versus low. By contrast, they should be more likely to accurately recall the relative difference when evaluability is low versus high.

Method

Design and Participants. Participants ($N = 400$, 197 females) recruited from MTurk were randomly assigned to conditions in a 2 (evaluability: low vs. high) $\times$ 2 (question type: absolute vs. relative difference) design.

Procedure. Participants were informed that the study aimed to understand how the general public perceives the Air Quality Index (AQI). They were presented with the following message: "The AQI is an index for reporting daily air quality. It tells you how clean or polluted your air is. The higher the AQI value, the greater the level of air pollution, and the greater the health concern." Participants in the high-evaluability condition were also provided with a chart about how to interpret AQI (see appendix), whereas such information was absent in the low-evaluability condition. Then, participants were told that the AQI of city A and city B is 120 and 180, respectively, and they were asked to judge how much worse the air quality of city B is than that of city A (1 = a little, 7 = a lot).

After performing some filler tasks that were unrelated to this study (about 5–10 minutes), participants were asked to recall how much the AQI difference was between the two cities that they had previously evaluated. Instead of using an open-ended question, we asked participants to choose between two options. In the absolute difference condition,

they were asked to choose between 50 and 60 (correct answer), whereas in the relative difference condition, the options were 50% (correct answer) and 60%.

Results and Discussion

Hypotheses Tests. Participants in the low-evaluability condition ($M = 4.78$, $SD = 1.33$) had much lower ratings for city B than for city A than those in the high-evaluability condition ($M = 4.33$, $SD = 1.32$; $F(1, 398) = 11.72$, $p = .001$, $d = .34$). The focal dependent variable here is accuracy rate. We conducted a logistic regression where evaluability, question type, and their interaction were included to predict the accuracy rate. This regression yielded two significant effects. First, the main effect of evaluability was significant (Wald $= 28.56$, $p < .001$), showing that participants were more likely to choose the right answer if the question was framed as the absolute difference (190 / 216 or 88%) rather than as a percentage (123 / 184 or 67%). More important, this main effect is qualified by its interaction with evaluability (Wald $= 9.38$, $p = .002$). To explore this interaction, we examined the effect of evaluability separately for each type of question. Specifically, when the question was framed as a percentage, the accuracy rate was higher in the low-evaluability (72 / 93 or 77%) than in the high-evaluability condition (51 / 91 or 56%; $\chi^2(1) = 9.48$, $p = .002$). The opposite was true when the question concerned the absolute difference: the accuracy rate was higher in the high-evaluability (104 / 114 or 91%) than in the low-evaluability condition (86 / 102 or 84%), though the difference did not reach significance ($\chi^2(1) = 2.43$, $p = .12$).

Discussion. Thus, collectively, the first three studies provide good evidence that evaluability indeed influences how individuals make numerical comparisons. Specifically, using three different approaches, these studies converge on the same notion that individuals are more likely to compute a ratio when they find the attribute values more difficult to evaluate. Building on this evidence, the subsequent studies examine the downstream effects on consumer preference and judgments.

STUDY 4

Study 4 was designed to test our theorizing by examining whether attribute evaluability can moderate previous findings that assumed reliance on absolute difference. Following previous research (Monga and Bagchi 2012; Pandelaere et al. 2011), we varied the absolute difference by framing the same relative difference using a less or more numerous unit (i.e., year vs. month). Participants were asked to choose from two options (see table 1): one option is cheaper ($849) but has a shorter warranty (5 years or 60 months), while another option is more expensive

TABLE 1

STIMULI USED IN STUDY 4

Choices	Year framing		Month framing	
	Warranty	Price	Warranty	Price
Option A	5 yrs.	$849	60 mo.	$849
Option B	6 yrs.	$899	72 mo.	$899

($899) but has a longer warranty (6 years or 72 months). Past research predicts that the choice share of the superior option (i.e., the one with a longer warranty) should be greater when the warranty is presented with month framing rather than year framing because the advantage looms larger under an expanded scale. As we discussed previously, these findings were based on the assumption that people tend to focus on the absolute difference. This is likely the case because attributes used in those studies are relatively familiar, and thus more evaluable. However, our theory suggests that this effect will be attenuated when evaluability decreases. To this end, we used two product categories: washing machine and dental crowns. Participants should perceive the warranty of a dental crown to be less evaluable than that of a washer because they are unlikely to know the distribution of possible values. Therefore, we expect to replicate previous findings for washers, but the same effect is less likely to be observed for dental crowns.

Method

Participants and Design. Participants ($N = 304$, 125 females) recruited from MTurk completed this 2 (framing: year vs. month) $\times$ 2 (evaluability: low vs. high) between-subjects study in exchange for a small monetary reward.

Procedure. Participants were told that the study examined consumer choice. In the low-evaluability (high-evaluability) condition, participants imagined choosing between two dental crowns (washers): a cheaper option with a shorter warranty and an expensive option with a longer warranty. These two attributes were crossed such that there is no clearly dominating option. The warranty was presented in either years or months, and the participants indicated their choice after reviewing the options. Participants then responded to a manipulation check of evaluability using the same measure used in study 1.

Results and Discussion

Manipulation Check. We conducted a 2 (framing: year vs. month) $\times$ 2 (evaluability: low vs. high) ANOVA to examine whether our manipulation of evaluability (operationalized as two categories) was successful. The analysis revealed only a main effect of product category ($F(1, 300) = 32.55$,

$p < .001$, $d = .65$). Specifically, participants rated the warranty information as more evaluable when the category was washer ($M = 3.06$, $SD = 2.05$) rather than dental crown ($M = 1.90$, $SD = 1.47$). The main effect of unit framing and the interaction effect did not reach significance ($ps > .80$).

Hypotheses Tests. To examine our predictions, we conducted a logistic regression where framing, evaluability, and their interaction were included to predict the choice share of the option with a longer warranty. This analysis revealed a marginally significant interaction only (Wald $\chi^2 = 3.51$, $p = .06$). No other effect was significant ($ps > .20$). To explore this interaction, we examined the conditional effect of framing on the choice share of the more expensive option. Specifically, we found that when evaluability was high (washer), we replicated our previous findings: participants in the month condition (32 / 68 or 47%) showed a stronger preference for the more expensive option than those in the year condition (22 / 72 or 31%; $\chi^2 = 4.02$, $p = .04$). However, we predicted that this effect would be attenuated in the dental crown condition, where low evaluability induces participants to compute the ratio (which was the same between the two framings). In support of our prediction, the likelihood of the more expensive dental crown being chosen did not significantly differ in the year (29 / 73 or 40%) and month (32 / 91 or 35%; $p > .50$) conditions.

Discussion. As a first test of our theory regarding how evaluability influences preference through shifting focus on absolute versus relative difference, study 4 shows that the effect of unit framing on preference reversal found in previous research is more likely to occur when the attribute values are easy to evaluate. By contrast, when individuals find it difficult to assess attribute values, the suppression of the effect suggests that, consistent with our predictions, participants were likely to calculate the ratio between values.

However, it is important to emphasize that although participants indeed perceived the warranty information of washers to be more evaluable than that of dental crowns, there might be other differences between these two product categories. Studies 5 and 6 address this limitation by manipulating evaluability while keeping the product category constant.

STUDY 5

Study 4 demonstrated that prior findings based on individuals' tendency to compute the absolute difference are mitigated when evaluability decreases. Following the same logic, study 5 was designed to test our theorizing by examining whether previous findings assuming participants' computing ratio (Hsee et al. 2009; Kwong and Wong 2006; Palmeira 2011) are attenuated when evaluability increases.

In addition, study 5 also aimed to evaluate an alternative interpretation. Specifically, although prior findings suggest that participants might have computed the ratio when evaluability is low, this tendency may also be caused by the size effect (Moyer and Landauer 1967) when people rely on the absolute differences. The size effect shows that for the same absolute distance, comparison is faster for smaller numbers (e.g., 3 and 5) than for larger numbers (e.g., 30 and 32), which implies that the difference will be perceived to be larger in the former than in the latter case. The literature has identified a number of explanations for this effect (Verguts, Fias, and Stevens 2005). For example, one possibility is that people encode numbers in a logarithmic manner such that their sensitivity to the same difference decreases as the number magnitude increases, and variations of this notion can be also found in other theories (Kahneman and Tversky 1979; Maglio, Trope, and Liberman 2013). This alternative account thus suggests that if participants compute the absolute difference (rather than the ratio), the same effect may still be observed.

The size effect appears to be a legitimate possibility. However, it remains unclear whether the size effect is sufficient to cause the preference reversal shown in previous studies. In other words, it might be an empirical question regarding which mechanism—the size effect or our proposed process—is the dominant driver. By testing the moderating role of evaluability, study 5 aims to separate this alternative account from our proposed mechanism because the size effect would not depend on evaluability, but our theory would predict so.

In addition, to further evaluate this alternative account, we included a third condition (i.e., zero condition) in addition to the small-ratio (.5 vs. .6; ratio = 1.2) and large-ratio (.1 and .2, ratio = 2) conditions (see table 2). In this new condition, the two attribute values are 0 and .3. When evaluability is low, we predicted that participants would be more likely to compute the ratio and therefore to choose the superior option in the large-ratio (small-ratio) condition. However, mathematically it is impossible to compute a ratio in the zero condition (Palmeira 2011). Therefore, the less evaluable attribute will remain unevaluable, pushing participants to rely on the other more evaluable attribute in this scenario—namely, price. Therefore, if our theorizing holds, the choice share of the superior option in the zero condition should be relatively low (Palmeira 2011). By contrast, the size effect would predict the opposite. Participants should be more sensitive to the difference between 0 and .3 than that between .1 and .2 and between .5 and .6. Therefore, if the observed difference between the small- and large-ratio conditions is caused by diminishing sensitivity, then the choice share of the superior option should be even higher, or at least not lower, in the zero condition than in other conditions.

TABLE 2

STIMULI USED IN STUDY 5

| | Conditions | | | | | |
| Choices | Small ratio | | Large ratio | | Zero | |
	Error	Price	Error	Price	Error	Price
Option A	.5	$39.99	.1	$39.99	0	$39.99
Option B	.6	$29.99	.2	$29.99	.3	$29.99

Method

Participants and Design. This study used a 3 (ratio: small vs. large vs. zero) × 2 (evaluability: low vs. high) between-subjects design. Participants ($N = 481$, 222 females) recruited from MTurk were randomly assigned to one of the conditions.

Procedure. Participants were asked to imagine that they were shopping for a digital thermometer and that they had narrowed their options to two products, A and B. The two options were identical except that option A was more precise and hence more expensive ($39.99) than option B ($29.99). The error margins of the two options were .5 °F and .6 °F in the small-ratio condition (ratio = 1.2), .1 °F and .2 °F in the large-ratio condition (ratio = 2), and 0 °F and .3 °F in the zero condition, where computing the ratio was impossible. In the high-evaluability condition, participants were informed that the typical error margins of home digital thermometers range from .1 °F to 1 °F, whereas no such information was provided in the low-evaluability condition.

Results and Discussion

Manipulation Check. We asked participants to indicate whether they had any idea about how good or bad the precision levels were using a seven-point scale ranging from 1 (I have no idea) to 7 (I have a clear idea). The results of a two-way ANOVA show that, in line with our expectations and prior literature (Yeung and Soman 2005), participants in the high-evaluability condition ($M = 3.88$, $SD = 2.22$) perceived the attribute values to be more evaluable than those in the low-evaluability condition ($M = 2.97$, $SD = 1.86$; $F(1, 475) = 23.45$, $p < .001$, $d = .44$). No other effect was significant ($ps > .85$).

Hypothesis Tests. We conducted a logistic regression where ratio manipulation, evaluability, and their interaction were included to predict the choice share of the more precise option. This analysis revealed three significant effects. The main effects of ratio manipulation (Wald $\chi^2 = 6.31$, $p = .012$) and evaluability (Wald $\chi^2 = 8.43$, $p = .004$) were both significant. More important, the interaction effect was significant (Wald $\chi^2 = 13.49$, $p < .001$).

We explored the implications of this interaction by focusing on low versus high evaluability separately.

In the low-evaluability condition, we replicated previous research (Hsee et al. 2009), showing that the choice share of the superior option is higher in the large-ratio (45 / 78 or 58%) than in the small-ratio (25 / 84 or 30%; $\chi^2(1) = 12.86$, $p < .01$) condition. In addition, we found that the likelihood of the superior option being chosen in the zero condition (32 / 85 or 38%) was lower than that in the large-ratio condition ($\chi^2(1) = 6.56$, $p = .01$) and not different from that in the small-ratio condition ($p = .28$). These results therefore provide support for our proposed mechanism, as the patterns are inconsistent with the predictions derived from the diminishing sensitivity account.

In the high-evaluability condition where range information was provided, we predicted that participants could make sense of the attribute values, leading to greater reliance on absolute differences. Thus, the difference between the small- and large-ratio conditions should be attenuated (because the absolute difference is the same). In support of our predictions, the choice share of the superior option did not differ between the small-ratio (22 / 78 or 28%) and large-ratio (28 / 74 or 38%; $p = .20$) conditions. However, this result might be attributed to an alternative account. That is, the evaluability manipulation (range information) might have trivialized this attribute, leading participants to rely on the price difference. If this was the case, then we should not observe any difference in the zero condition. However, our theory would predict something different. Specifically, we propose that if the range information enabled participants to make sense of the absolute difference, their preference for the superior option should be elevated in the zero condition in which the absolute difference is larger than the other two conditions (i.e., .3 vs. .1). In line with our reasoning, we found that the choice likelihood of the superior option was significantly higher in the zero condition (48 / 82 or 59%) than in the small-ratio ($\chi^2(1) = 14.94$, $p < .001$) or large-ratio ($\chi^2(1) = 6.67$, $p < .01$) condition.

Discussion. Earlier research showing ratio effects might have deliberately employed less evaluable stimuli, implying that the ratio effect should be less likely to arise if the attributes are easy to evaluate. Nonetheless, this implication has never been empirically tested. Examining this possibility, study 5 shows that the ratio effect is indeed moderated by attribute evaluability. The results demonstrate that the ratio effect that we observed in the low-evaluability condition was suppressed when participants received additional range information that helped them make sense of attribute values.

By demonstrating the moderating role of evaluability, this study also lends additional support to our theorizing because, as discussed previously, the alternative explanation would not predict this interaction effect. In addition,

the results observed in the zero condition further suggest that previous findings are unlikely to be solely attributable to the size effect.

STUDY 6

Study 6 had several objectives. Foremost, although studies 1–3 jointly showed that evaluability influences people's tendency to compute the difference versus ratio, it remains to be tested whether the preference reversals observed in studies 4 and 5 were indeed caused by such tendencies. Study 6 was thus designed to evaluate the underlying mechanism by using a mediation approach. In addition, while studies 4 and 5 used choice as dependent measure, study 6 attempted to test the robustness of our findings by using continuous judgments. To this end, we opted for a context that also contains practical implications. Specifically, we told participants how many grams of saturated fat two cake options have and asked them to judge, relative to the healthier one (which has less saturated fat), how much unhealthier the other cake is (using a relative judgment makes the results comparable across conditions). Similar judgments appear to be quite ubiquitous and frequent in consumer lives. Our theorizing suggests that consumers may draw very different or even opposing implications from nutrition information as a function of evaluability. While in studies 4 and 5, we show that the effects were suppressed when evaluability increases or decreases, study 6 tested whether the pattern could be flipped. To this end, we created two pairs of numbers (see table 3). Cake A and cake B had 2 grams and 12 grams of fat in condition 1 but 40 grams and 80 grams of fat in condition 2. Condition 1 is higher than condition 2 in terms of relative difference (6 times vs. 2 times) but lower than condition 2 in terms of absolute difference (10 grams vs. 40 grams). If individuals rely on the absolute difference, then their perceived unhealthiness of cake B should be higher in condition 2 than in condition 1. However, if individuals rely on the relative difference, then the opposite is expected: their perceived unhealthiness of cake B should be higher in condition 1 than in condition 2.

Method

Design and Participants. Undergraduate participants ($N = 228$, 131 females) were randomly assigned to conditions in a 2 (condition 1: 2–12 vs. condition 2: 40–80) × 2 (evaluability: low vs. high) between-subjects design.

Procedure. The study was described as a food healthiness perception study. Participants imagined that there are two pieces of cake with a different amount of saturated fat. In the high-evaluability condition, participants were also informed that "the Dietary Guidelines for Americans recommend consuming less than 20 grams of saturated fat per

TABLE 3

STIMULI USED IN STUDY 6

	Conditions	
	Condition 1	Condition 2
Absolute difference	Low (10g)	High (40g)
Relative difference	High (6×)	Low (2×)
Cake A	2g	40g
Cake B	12g	80g

day," whereas those in the low-evaluability condition did not receive such information. All participants were instructed to indicate, relative to cake A, how much unhealthier cake B is using a seven-point scale (1 = a little, 7 = a lot). To assess the underlying process, we asked participants how they responded to the previous question. Specifically, we presented the following message: "How did you respond to the previous question? We asked some participants this question, and here is what we got. Some people said that they did subtraction: that is, computing the difference between the two numbers. Some people said that they did division: that is, computing the ratio between the two numbers." After reading this paragraph, all participants indicated which calculation was closer to what they actually did mentally in the previous judgment using a seven-point scale (1 = I did subtraction; 7 = I did division).

Results and Discussion

Unhealthiness Judgment. The results of a 2 × 2 ANOVA on participants' perceived unhealthiness of cake B yielded a significant interaction effect ($F(1, 224) = 9.75$, $p = .002$, $d = .42$) only. Other effects were not significant (ps > .50). We conducted contrast analyses to understand the meaning of the interaction effect. In the low-evaluability condition, the cake containing more fat was judged as relatively unhealthier in condition 1 ($M = 5.76$, SD = 1.32) than in condition 2 ($M = 4.88$, SD = 1.95; $t(224) = 2.50$, $p = .01$, $d = .53$). Because the relative (absolute) difference was larger (smaller) in condition 1 than in condition 2, this result suggests participants' greater reliance on relative difference. By contrast, in the high-evaluability condition where the daily reference value presumably would help participants make sense of the numbers, the reverse was true. In this case, cake B was rated as unhealthier in condition 2 ($M = 5.68$, SD = 1.83) than in condition 1 ($M = 5.10$, SD = 1.82; $t(224) = 1.89$, $p = .06$, $d = .32$), suggesting participants' reliance on the absolute difference.

Type of Calculation. We have proposed that evaluability shifts individuals' tendency to compute the absolute or relative difference. A 2 × 2 ANOVA on calculation tendency yielded a significant main effect of evaluability only ($F(1, 224) = 6.29$, $p = .01$, $d = .34$). Supporting our

FIGURE 1

MEDIATION MODEL (STUDY 6)
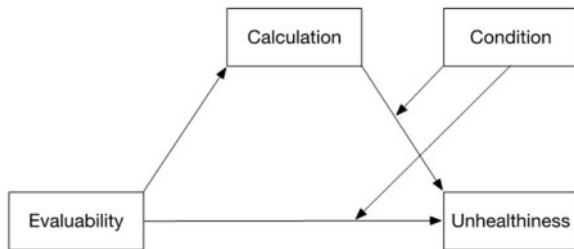
reasoning and replicating the findings from studies 1–3, participants in the low-evaluability condition ($M = 4.46$, $SD = 2.01$) had a stronger tendency to compute the ratio than those in the high-evaluability condition ($M = 3.73$, $SD = 2.29$).

Mediation Analysis. Finally, we assessed the underlying role of calculation type in the impact of evaluability, using the PROCESS macro (model 15; figure 1) developed by Hayes (2013). This moderated mediation analysis yielded supportive evidence for our theorizing. Although the interaction between condition and evaluability remained significant ($t(222) = 2.74$, $p = .01$) when the mediator was included in the model (the condition × mediator interaction was significant: $t(222) = 2.42$, $p = .02$), the results of bootstrapping with 5,000 samples revealed that the indirect effect (moderated mediation) was significant, with a 95% confidence interval excluding zero (.03, .52).

Discussion. By demonstrating the downstream effect of our theorizing on unhealthiness judgments, this final study shows the generalizability and robustness of our conceptual framework. In addition, this study replicated the findings from studies 1–3 that people are more likely to compute the relative difference when evaluability decreases. More importantly, this study provides further evidence for the underlying mechanism by showing that the difference in the tendency to compute the absolute versus relative difference mediates the effect of evaluability on judgments.

GENERAL DISCUSSION

Consumers typically rely on either the absolute difference or relative difference to form their comparative judgment. While previous research has largely focused on the influence of either absolute or relative difference, we draw on prior literature on evaluability to provide an integrated framework. Specifically, we propose that consumers are more likely to compute and rely on the absolute difference when attribute values are easy to evaluate (that is, when

consumers are aware of the reference point). By contrast, when attribute values are more difficult to evaluate (that is, when consumers cannot make sense of the absolute difference), consumers tend to compute the relative difference, which is more interpretable. Six studies provide triangulating support for this theorizing by using multiple operationalizations of evaluability (e.g., reference point, range information) and different product categories (e.g., washer, yoga mat, digital thermometer, and cake).

The current conceptualization and findings contribute to the consumer literature in multiple ways. First, our conceptualization presents a new perspective on how people make numerical comparative judgment, delineating conditions under which the assessment is more likely to be driven by absolute or relative differences. Second, and relatedly, we also identify boundary conditions for two sets of findings in the literature. Specifically, we show that previous findings that assumed reliance on absolute difference are attenuated when attribute values become difficult to evaluate. In contrast, preference change based on consumer reliance on relative difference is suppressed when attribute values become more evaluable.

In addition, the present work not only draws on research on evaluability (Hsee 1996; Hsee et al. 1999) but also adds to this literature in two ways. On the one hand, while past research has shown that evaluability can lead to preference reversal when the decision mode changes from a separate judgment to a joint evaluation, we show that under situations in which the evaluation mode is kept constant (joint evaluations in our context), attribute evaluability can still systematically influence consumers' perceptions and preferences through a very different mechanism. On the other hand, previous findings imply that individuals' reliance on difficult-to-evaluate attributes increases when the evaluation mode changes from separate to joint. Nevertheless, how such information is interpreted remains unclear. While participants have a clear idea that having written 70 programs is better than having written 10, in deciding how much better one candidate is than another, they still need information about "which value is the best possible," "which value is the worst possible," and "what the value distribution of the attribute is" (Hsee et al. 1999). Our hypothesis and findings provide one possible answer to this question by suggesting that when attribute values are difficult to comprehend, consumers may seek to make sense of these numbers by computing the relative ratio.

Finally, the present findings may also have implications for empirical modeling. Specifically, the utility that consumers derive from a product could be modeled as a function of the raw (x, y) or log-transformed attribute values ($\log x$ or $\log y$). When the raw attribute values are used, the utility difference derived from these two products would be a function of $x - y$ (absolute difference). In contrast, when log transformation is used, the utility difference derived from these two products would be a function of x / y

because $\log x - \log y = \log (x / y)$ (relative difference). While whether to use raw or log-transformed attribute values might depend on many factors, the current theory suggests that attribute evaluability is one additional factor to consider. We leave to future research the task of empirically evaluating whether taking evaluability into account (and using raw or log-transformed attribute values accordingly) would yield better performance.

Apart from advancing theoretical knowledge, our work also provides applied implications for marketing practitioners. For example, price discounts could be framed as absolute difference (e.g., $50 off) or relative difference (e.g., 30% off). The question then arises as to whether consumers would respond to these two forms differently, and if yes, which format marketers should use. The present findings suggest that communicating discounts in absolute form might be better for highly evaluable products such as groceries because consumers often have a highly accessible reference point. In contrast, the discount should be presented in relative difference when consumers are not very familiar with the prices. Some anecdotal observations are in line with this reasoning. For example, high-end stores such as Nordstrom and Godiva often use percent form, whereas places like Walmart and Hershey's often communicate discounts using absolute dollar amount.

In addition, the current conceptualization also has some implications for brand naming strategies. One common practice for naming products is to use alphanumeric names, which combine numbers and letters. This practice is especially pervasive in technological product categories—for instance, headphones (e.g., Bose-QuietComfort 35, Sony 1000X), digital cameras (e.g., Canon EOS 80D, Nikon D5600), computers (e.g., Dell Latitude 7200, HP Pavilion x360), cars (e.g., Audi A4, Lexus RX350), drones (e.g., DJI Phantom 4, GoPro-HERO6), and routers (e.g., Linksys EA9300, Motorola N450). Given that many consumers may not have sufficient knowledge about these products (i.e., low attribute evaluability), they may simply rely on the relative difference in brand names to infer performance difference. For example, consumers may perceive an air purifier named RX1000 as twice effective as a model named RX500. Such inference is less likely if the two models are named RX6000 and RX5500.

DATA COLLECTION INFORMATION

The author supervised the collection of data for studies 1 and 6 by research assistants at the University of Texas at San Antonio behavioral lab between spring 2016 and spring 2017. The author collected data for studies 2–5 using Amazon Mechanical Turk in summer 2017 (studies 2 and 3), fall 2015 (study 4), and spring 2018 (study 5).

APPENDIX: CHART USED IN THE HIGH-EVALUABILITY CONDITION (STUDY 3)

Category, Color & Range	What does this mean?	Precautions to take
Good 0-50	Air quality is good.	None: Everyone enjoy outdoor activities.
Moderate 51-100	Air quality is a concern for people who are extra sensitive to air pollution.	People extra sensitive to air pollution: Plan strenuous outdoor activities when air quality is better.
Unhealthy for Sensitive Groups 101-150	Air quality is unhealthy for many people including active adults, people with lung disease (including asthma), older adults and children.	Sensitive groups: Cut back or reschedule strenuous outdoor activities.
Unhealthy 151-200	Air quality is unhealthy for everyone, especially people with heart or lung disease.	Everyone: Avoid strenuous outdoor activities.
Very Unhealthy 201-300	Air quality is unhealthy for everyone, especially people with heart or lung disease.	Everyone: Avoid physical outdoor activities.
Hazardous 301-500	Air quality is hazardous for everyone.	Everyone: Avoid all outdoor activities.

REFERENCES

Adaval, Rashmi (2013), "Numerosity and Consumer Behavior," *Journal of Consumer Research*, 39 (5), xi–xiv.

Adaval, Rashmi and Kent B. Monroe (2002), "Automatic Construction and Use of Contextual Information for Product and Price Evaluations," *Journal of Consumer Research*, 28 (4), 572–88.

Biswas, Abhijit, Sandeep Bhowmick, Abhijit Guha, and Dhruv Grewal (2013), "Consumer Evaluations of Sale Prices: Role of the Subtraction Principle," *Journal of Marketing*, 77 (4), 49–66.

Burson, Katherine A., Richard P. Larrick, and John G. Lynch Jr. (2009), "Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals," *Psychological Science*, 20 (9), 1074–8.

Gaston-Breton, Charlotte (2006), "The Euro Impact on the Consumer Decision Process: Theoretical Explanation and Empirical Evidence," *Journal of Product & Brand Management*, 15 (4), 272–9.

Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, New York: Guilford.

Hsee, Christopher K. (1996), "The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes*, 67 (3), 247–57.

———. (2000), "Attribute Evaluability and Its Implications for Joint-Separate Evaluation Reversals and Beyond," in *Choices, Values, and Frames*, ed. Daniel Kahneman and Amos Tversky, New York: Cambridge University Press, 543–63.

Hsee, Christopher K., George F. Loewenstein, Sally Blount, and Max H. Bazerman (1999), "Preference Reversals between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis," *Psychological Bulletin*, 125 (5), 576–90.

Hsee, Christopher K., Yang Yang, Yangjie Gu, and Chen Jie (2009), "Specification Seeking: How Product Specifications Influence Consumer Preference," *Journal of Consumer Research*, 35 (6), 952–66.

Hsee, Christopher K. and Jiao Zhang (2010), "General Evaluability Theory," *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 5 (4), 343–55.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decisions under Risk," *Econometrica*, 47 (2), 263–92.

Krider, Robert E., Priya Raghubir, and Aradhna Krishna (2001), "Pizzas: Π or Square? Psychophysical Biases in Area Comparisons," *Marketing Science*, 20 (4), 405–25.

Kwong, Jessica Y. Y. and Ellick K. F. Wong (2006), "The Role of Ratio Differences in the Framing of Numerical Information," *International Journal of Research in Marketing*, 23 (4), 385–94.

Maglio, Sam J., Yaacov Trope, and Nira Liberman (2013), "Distance from a Distance: Psychological Distance Reduces Sensitivity to Any Further Psychological Distance," *Journal of Experimental Psychology: General*, 142 (3), 644–57.

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.

Moyer, Robert S. and Thomas K. Landauer (1967), "Time Required for Judgements of Numerical Inequality," *Nature*, 215 (5109), 1519–20.

Palmeira, Mauricio M. (2011), "The Zero-Comparison Effect," *Journal of Consumer Research*, 38 (1), 16–26.

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.

Pope, Devin and Uri Simonsohn (2011), "Round Numbers as Goals: Evidence from Baseball, Sat Takers, and the Lab," *Psychological Science*, 22 (1), 71–9.

Shafir, Eldar B., Daniel N. Osherson, and Edward E. Smith (1993), "The Advantage Model: A Comparative Theory of Evaluation and Choice under Risk," *Organizational Behavior and Human Decision Processes*, 55 (3), 325–78.

Verguts, Tom, Wim Fias, and Michaël Stevens (2005), "A Model of Exact Small-Number Representation," *Psychonomic Bulletin & Review*, 12 (1), 66–80.

Wertenbroch, Klaus, Dilip Soman, and Amitava Chattopadhyay (2007), "On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects," *Journal of Consumer Research*, 34 (1), 1–10.

Wong, Kin Fai Ellick and Jessica Y. Y. Kwong (2005), "Comparing Two Tiny Giants or Two Huge Dwarfs? Preference Reversals Owing to Number Size Framing," *Organizational Behavior and Human Decision Processes*, 98 (1), 54–65.

Wright, John H. (2001), "The Influence of Absolute Differences and Relative Differences on Unidimensional Difference Judgments," doctoral dissertation, Marketing Department, University of Chicago, Chicago, IL 60637.

Yeung, Catherine W. M. and Dilip Soman (2005), "Attribute Evaluability and the Range Effect," *Journal of Consumer Research*, 32 (3), 363–369.