Cognition 194 (2020) 104043

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Gamble evaluation and evoked reference sets: Why adding a small loss to a gamble increases its attractiveness

Craig R.M. McKenzie[a,*], Shlomi Sher[b]

[a] UC San Diego, United States
[b] Pomona College, United States

ARTICLE INFO

Keywords:
Choice
Judgment
Norms
Rationality
Framing

ABSTRACT

When presented with a gamble involving a chance of winning $9, participants rate it as only moderately attractive. However, when other participants are presented with a gamble that adds a chance of losing 5 cents – resulting in gamble that is strictly worse – they rate it as much more attractive. This surprising effect has previously been explained in terms of the small loss increasing the affective evaluability of $9. This paper argues for an alternative model, in which the baseline and small-loss gambles evoke different reference sets for comparison. In inferring a relevant reference set, people are sensitive to both the objective content and the framing of a gamble. The model distinguishes between two effects of evoked reference sets on behavior – an obligatory (and rational) effect on scale interpretation, and an optional (but not rational) effect on the internal representation of value. Five experiments provide strong evidence for the evoked reference set model. Data from attractiveness ratings suggest large and consistent reference set effects on scale interpretation, while data from willingness-to-pay and choice tasks suggest that effects on the internal representation of value are less robust.

1. Introduction

Context effects in decision making are said to occur when seemingly irrelevant changes to the choice environment affect judgments and choices (e.g., Huber, Payne, & Puto, 1982; Lichtenstein & Slovic, 1971; Payne, Bettman, & Johnson, 1993; Read, Olivola, & Hardisty, 2017). Such effects are of interest in part because they appear to violate principles of rational choice. However, it is often difficult to know whether a change in context is irrelevant or instead provides information that a rational actor would utilize. Thus, it is crucial to analyze what information might be conveyed by a given change in context, as well as whether the information is sufficient to explain participants' behavior (e.g., McKenzie, Sher, Leong, & Müller-Trede, 2018; Sher & McKenzie, 2006, 2011, 2014; Wernerfelt, 1995).

Slovic, Finucane, Peters, and MacGregor (2002) reported an intriguing effect in risky choice. Some participants were presented with a "standard" gamble with a 7/36 chance of winning $9 and a 29/36 chance of winning nothing. How attractive would it be to play the gamble one time? Participants' mean response was 9.4 on a scale ranging from 0 (not at all attractive) to 20 (very attractive). Other participants were presented with the same gamble, but a small loss component was added. Instead of a 29/36 chance of winning nothing, there was a 29/36 chance of losing 5 cents. Although the small-loss gamble is strictly worse than the standard gamble, the mean attractiveness rating increased to 14.9. Adding the small loss affected choices, as well. Compared to those presented with the standard gamble, participants presented with the small-loss gamble were more likely to prefer playing the gamble to receiving $2 for sure (61% vs. 33%).

To explain these seemingly counter-normative effects, Slovic et al. (2002) hypothesized that participants use an "affect heuristic", which assumes that, "in the process of making a judgment or decision, people consult or refer to the positive and negative feelings consciously or unconsciously associated with the mental representations of the task" (Bateman, Dent, Peters, Slovic, & Starmer, 2007, p. 366). Slovic et al. proposed that, without the small loss outcome, it is difficult to know how good the $9 outcome is, since there is nothing to compare it with. Because it is difficult to evaluate $9 on its own, participants focus on the 7/36 probability, which is low, and therefore rate the gamble as only slightly attractive. By adding the small loss outcome, however, $9 becomes "evaluable" (e.g., Hsee, 1996) and "comes alive with feeling" (Bateman et al., 2007), thereby increasing the gamble's appeal. We will refer to this explanation as the "affective evaluability" account.

While the affective evaluability account is plausible, it fails to address some basic questions. First, why does "lose 5 cents" increase the affective evaluability of $9 more than "win nothing" does? Perhaps there is something special about a $0 outcome that hinders evaluability

* Corresponding author at: UC San Diego, 9500 Gilman Drive – MC 0553, La Jolla, CA 92093-0553, United States.
E-mail address: cmckenzie@ucsd.edu (C.R.M. McKenzie).

https://doi.org/10.1016/j.cognition.2019.104043
Received 22 August 2016; Received in revised form 6 August 2019; Accepted 8 August 2019

(e.g., it cannot enter into ratio calculations). However, Bateman et al. (2007) reported that a gamble with the second outcome described as "lose nothing" was rated as much more attractive than the same gamble with the second outcome described as "win nothing". The affective evaluability of the $9 outcome appears to be the same in the two $0-outcome gambles, so why are the resulting ratings so different? Bateman et al. suggested that the difference was due to the more "positive tone" of "lose nothing", but this explanation is different from the affective evaluability explanation (because evaluability is no longer a component). Parsimony is reduced when one account (affective evaluability) is used to explain the difference in ratings when the second outcome is "win 0" vs. "lose 5 cents" and a second account ("tone") is used to explain the difference when the outcomes are "win 0" vs. "lose 0". Finally, Bateman et al. showed that adding a small *gain* outcome of 5 cents also increased the attractiveness of the standard gamble, but not as much as adding the small loss did. Why is the evaluability of $9 aided more by the small loss than by the small gain?

1.1. Evoked reference sets

An alternative and more parsimonious explanation of the results is that the different gambles evoke different reference sets for comparison. When presented with a single gamble and asked how attractive it is, the natural question is: Compared to what? Indeed, this question must be answered by participants at some level in order to interpret the subjective attractiveness scale (e.g., what counts as a 12 on the scale?). Because the standard gamble involves only wins, we suspect that, for many participants, the evoked reference set consists of other gambles involving only wins. However, when the small loss is added, the resulting gamble includes both a winning and a losing outcome, and will presumably evoke a reference set of gambles involving wins and losses. Relative to an inferior reference set, any fixed gamble will be more attractive by comparison. Thus, when the small loss is added, the downward-shifted reference set will tend to inflate the gamble's rated attractiveness. In particular, relative to a win-lose reference set, a 5-cent loss is about as good as a loss can be, while, relative to a win-win reference set, winning nothing is the worst possible outcome. The same account naturally explains why the gamble is rated as much more attractive when the second outcome is described as "lose 0" (suggesting a win-lose reference set) rather than "win 0" (suggesting a win-win set), although all outcomes are the same.

Note that, while the affective evaluability account focuses on the difficulty of putting the *$9 outcome* in context, the evoked reference set account focuses on how to put the *entire gamble* in context. We are suggesting that the small-loss effect on attractiveness ratings is due to the gambles being compared to the different reference sets that their outcomes or descriptions evoke.

Much research has shown that subjective ratings can be influenced by context, especially by other stimuli presented during the experiment, which serve as the reference set (e.g., Parducci, 1965, 1995). Importantly, though, Slovic et al. (2002) presented only a single gamble for participants to rate; that is, there were no other gambles presented to serve as a reference set. However, there is evidence that, when a target stimulus belongs to a salient category, people will appropriate the category as a reference set. For example, when rating the heights of a series of males and females on a subjective scale, participants use the target's gender as a reference set (Biernat, Manis, & Nelson, 1991). That is, if the target is female, participants will rate the target's height relative to other females. Thus, the same objective height, say 5′ 7″, could be rated as "somewhat tall" if the target is female, but as "somewhat short" if the target is male.

The standard and small-loss gambles studied by Slovic et al. (2002) differ in terms of whether only wins are involved, or both wins and losses, and it has been shown that the inclusion of positive and/or negative outcomes can influence behavior in a variety of ways. For example, the presence of a potential loss in risky choice increases attention and physiological arousal (Yechiam & Hochman, 2013). Reference sets are also influenced by whether outcomes are positive or negative. McGraw, Larsen, Kahneman, and Schkade (2010) noted that loss aversion ("losses loom larger than gains") is robust in choice tasks but often disappears when participants rate how gains and losses would make them feel. They argued that, while choice compels participants to directly compare gains and losses, asking for ratings often leads participants to evaluate the outcome relative to other outcomes with the same valence. That is, when asked to rate the intensity of losing $15, it is natural to rate it relative to other losses, and when asked to rate a gain of $15, it is natural to compare it with other gains. Because they are normalized to different implicit reference sets, comparing such ratings can be misleading. In another study, following up on an experiment by Birnbaum (1999), Leong, McKenzie, Sher, and Müller-Trede (2019) elicited subjective ratings of the "largeness" of the numbers 2 and -2 in a between-subjects design. On average, participants rated the number 2 as smaller than the number -2, because, those authors argued, 2 evokes a reference set of natural numbers (relative to which 2 is small) while -2 evokes a reference set of integers (relative to which -2 is neither large nor small).[1]

In a theoretical article, Kahneman and Miller (1986) proposed a similar explanation of the (then unpublished) small-loss effect on attractiveness ratings. They suggested that the standard gamble and the small-loss gamble evoke different norms (an evoked set of exemplars that serves as a representation of a normal outcome or category member, roughly analogous to a reference set) because the gambles differ in terms of "the presence or absence of risk of loss" (p. 142). "The [small-loss] bet appears very favorable among bets that involve a risk of loss, but a modest chance to win $9 is mediocre in a context of purely positive prospects" (pp. 141–142). Our account can be viewed as both a refinement and an extension of their explanation and, for the first time, we present empirical evidence in support of it. In what follows, we distinguish between two different routes whereby reference sets can affect ratings and choices, and we relate the reference sets that gamble descriptions evoke to task-relevant information that "leaks" from a speaker's choice of frame (Sher & McKenzie, 2006).

Fig. 1 illustrates our proposed account of how a single gamble is evaluated in isolation. We focus first on the process involved in rating a gamble on an ambiguous subjective scale (choices are considered below). To generate a rating, it is necessary to form some internal representation of the value of the gamble, and also to construct an interpretation of the rating scale, in order to map the internally represented value onto a specific scale level. That is, the subjective appeal of the gamble must be determined, and this internal representation of value must be translated into an informative response on the rating scale. We assume that a gamble is composed of not only its objective components (probabilities and outcomes), but also its description or framing (e.g., "win 0" vs. "lose 0"). The objective components of the gamble directly influence the internal representation of value, independent of the specific scale used to report evaluations. Highly probable large gains make a gamble valuable, for example.

To interpret an ambiguous rating scale (e.g., a 0–20 "attractiveness" scale), a reference set must be used to anchor the scale values. Holding the internal representation of subjective value fixed, the better the reference set is, the lower the rating corresponding to that level of subjective value will be. For example, a gamble which offers a chance of

[1] In Birnbaum's (1999) original study, participants in a between-subjects design rated the number 9 as larger than the number 221. Consistent with the present analysis, Birnbaum speculated that 9 evokes a reference set of one-digit numbers whereas 221 evoked a reference set of three-digit numbers. However, Leong et al. (2019) showed that the "9 > 221 effect" is more complicated: It critically depends on the numerical values included in the rating scale itself, which, in this task, may also influence the evoked reference set to which the target is compared.

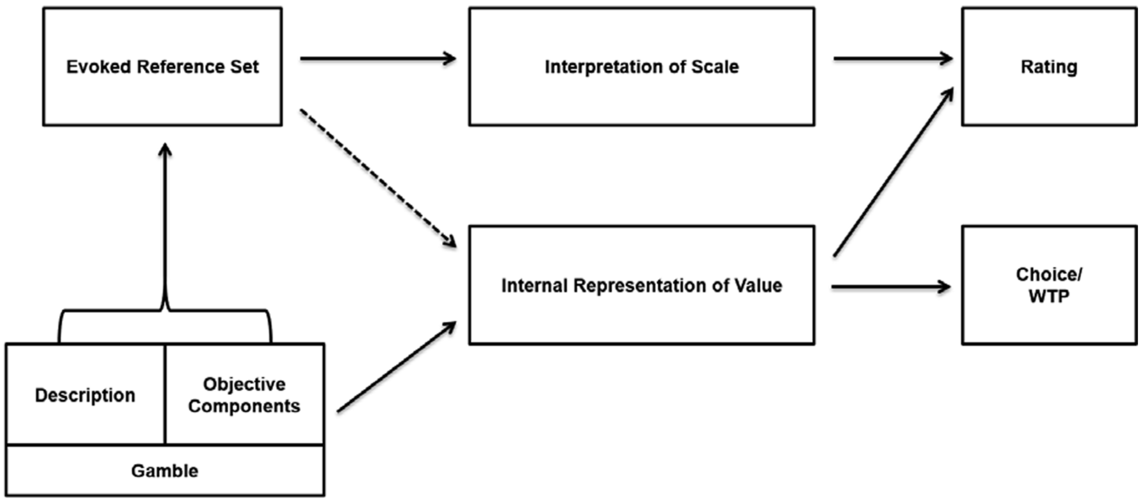

Fig. 1. Proposed model of how a single gamble is evaluated. The objective components of the gamble affect its internal representation of value, and the objective components and the gamble's description (or frame) evoke a reference set. When the gamble is evaluated on an underspecified subjective scale, the reference set must be used to interpret the scale (top solid arrow), and an inferior evoked reference set leads to higher ratings. The model includes a second optional path (broken arrow) whereby reference sets may directly induce contrast effects on internal representation of value, potentially leading to scale-independent effects.

winning $9 will receive a much lower rating if the top rating of 20 corresponds to a chance to win $1000 rather than a chance to win $10. In Fig. 1, the objective components of the gamble not only influence the internal representation of value, but also influence which reference set is evoked. Importantly, when only a single option is presented, the relevant range of outcomes is unspecified. What should participants do under such circumstances? One strategy would be for them to throw up their hands and respond either randomly or with the midpoint on the scale. At least some participants do not adopt this strategy, as different gambles lead to systematically different responses. Instead, participants apparently try to provide meaningful responses by answering the implicit "Compared to what?" reference set question as best they can. In particular, we assume that the presentation of a gamble involving only wins is relatively likely to evoke a reference set of other gambles involving only wins, whereas a gamble with potential win and loss outcomes is more likely to evoke a reference set of gambles involving both wins and losses.

This pattern of behavior is reasonable if participants assume that the focal gamble has been drawn from a larger relevant consideration set. For example, a gamble involving both wins and losses could not have been sampled from a population of gambles involving only wins. A single gamble can provide compelling evidence regarding the population from which it was sampled. Furthermore, human decision makers are highly sensitive to such evidence, drawing strong population inferences from small samples when little other information is available (Sher & McKenzie, 2014; see also Vul, Goodman, Griffiths, & Tenenbaum, 2014).

We also assume that the description (or framing) of the gamble can influence the evoked reference set. A gamble with outcomes of "win $9" and "win $0" and a gamble with outcomes of "win $9" and "lose $0" have the same objective outcomes, but different descriptions. The former description may be more likely to evoke a reference set of gambles involving only wins, and the latter a reference set involving wins and losses. Whether this is reasonable depends on whether the *description* provides evidence about the reference set (context) from which the single gamble was sampled. If "speakers" are more likely to describe a $0 outcome as "win nothing" when other gambles in the immediate context involve only wins compared to when the other gambles involve wins and losses, then it would be reasonable for a "listener" to infer that other gambles in the set involve only wins when the $0 outcome is framed as "win nothing" (see Leong, McKenzie, Sher, & Müller-Trede, 2017; McKenzie & Nelson, 2003; Sher & McKenzie, 2006; Teigen & Karevold, 2005). Below, we report evidence that people are, in fact, more likely to describe a $0 outcome as "win nothing" (rather than "lose nothing") when other gambles in the set involve primarily wins rather than a mix of wins and losses. Thus a gamble with a small loss, or with a zero outcome framed as a loss, suggests an inferior reference set for the anchoring of scale values, resulting in an increase in ratings.

The preceding description of the rating process is essentially a simple rational analysis of the small-loss effect on attractiveness ratings. The effect can be explained by the obligatory use of an implicit context to interpret an ambiguous judgment scale, assuming that participants draw reasonable inferences about underlying reference sets from the (framed) gamble they sample. However, in Fig. 1, the broken arrow from "evoked context" to "internal representation of value" represents another potential route in the model that, though neither obligatory nor normative, is psychologically plausible. Evoked contexts might influence ratings not just via scale interpretation, but also via effects on the internal representation of value. It is possible, for example, that adding the small loss not only increases ratings because an inferior reference set is used to interpret the scale, but also because the inferior reference set makes the gamble "feel" more valuable by subjective contrast.

How can we know whether attractiveness ratings are being driven by scale interpretation only, or by both scale interpretation and changes in internal representation of value? One way is to eliminate scale interpretation as a feasible route and see if the small-loss effect persists. If it does, then this would indicate that evoked context is influencing the internal representation of value over and above its effect on scale interpretation. This possibility is most naturally examined by using choices rather than subjective ratings as the dependent measure.

1.2. Ratings vs. choices

Recall that Slovic et al. (2002) reported that, when given a choice between playing the gamble one time and receiving $2 for sure, participants more often preferred the gamble when the 5-cent loss was added. That is, participants not only rated the small-loss gamble as more attractive, but also more often preferred it to a sure amount. Because there is no ambiguous rating scale to interpret, no reference set needs, in principle, to be evoked in this choice task. Instead, the two options, the gamble and a certain $2, only need to be compared to each other. Whether the gamble was drawn from a set involving only wins or from a set involving wins and losses has no obvious normative relevance for preference between the gamble and a sure gain of $2.

Choice effects can nonetheless be explained by the model in Fig. 1 if reference sets (a) are evoked even in choice tasks and (b) influence the internal representation of value (broken arrow). Specifically, the perceived value of the gamble may exhibit a contrast effect relative to the evoked reference set, making it feel more choiceworthy when the reference set is inferior. Such an effect could be explained by multi-attribute choice models (e.g., Mellers & Cooke, 1994) which apply the principles of range-frequency theory (Parducci, 1965) to account for context effects on perceived value. It would also be consistent with decision by sampling theory (Stewart, Chater, & Brown, 2006; Stewart, Reimers, & Harris, 2015; but see Alempaki et al., in press), according to which the subjective value of a particular gain (or loss) is determined by a series of binary ordinal comparisons with other gains (or losses) in memory. If a zero outcome framed as a "win" is compared to other remembered wins, the "win 0" outcome will fare poorly in the relvant binary comparisons while the "lose 0" and "lose 5 cent" outcomes will fare well relative to remembered losses. We revisit the relationship between our proposed account and decision by sampling theory in the General Discussion.

Because choice tasks eliminate the need for response scale interpretation, the model in Fig. 1 implies that, if there is any effect on choice of adding the small loss to the standard gamble, then the evoked reference set is influencing the internal representation of value. The same reasoning applies to willingness-to-pay (WTP) judgments, for which the response scale has a definite objective meaning independent of the reference set. Whereas there are two routes in the model for evoked reference sets to influence subjective ratings, there is only one route to influence choice and WTP. By examining the pattern of results in judgment vs. choice tasks, we should be able to draw conclusions about the relative influence of evoked reference sets on the internal representation of value in both tasks. For example, if the small-loss effect is much weaker, or non-existent, for choices as compared to ratings, this would suggest not only that evoked reference sets have little or no influence on choice, but also that scale interpretation is accounting for most or all of the effect in the ratings task. However, if the small-loss effect is similar for both ratings and choices, this would suggest that evoked reference sets influence perceived value in both tasks, and that differences in scale interpretation are not essential for the small-loss effect on ratings.

Bateman et al. (2007) and Slovic et al. (2002) reported a total of four choice experiments examining the small-loss effect. They looked at participants' country of origin (US and UK) and the size of the sure amount that was offered along with the gamble (2 dollars/pounds and 4 dollars/pounds). The two experiments using 2 dollars/pounds as the sure amount resulted in a significant effect: Participants preferred the gamble to the sure amount more often when the small loss was added to the standard gamble. The two experiments using 4 dollars/pounds did not result in a significant effect. It might be that the larger sure amount resulted in a floor effect and masked the small-loss effect. However, it could also be that the effect on choices is less robust than the effect on ratings. More data are needed to see if the small-loss effect on choice is as strong and consistent as the effect on ratings.

1.3. Overview of experiments

In what follows, we report five experiments that replicate the small-loss effect on ratings, test competing predictions of the affective evaluability and evoked reference set accounts, and address the two routes, depicted in Fig. 1, whereby evoked contexts can influence behavior. Experiment 1 replicates and extends earlier research by comparing judgments and choices for the three critical gambles (in which the second outcome is "win 0", "lose 5 cents", or "lose 0") within a single design. Experiment 2 provides a critical test of the competing accounts by not only manipulating the small outcome, but also manipulating whether the large outcome is a $9 win or a $9 loss. The affective evaluability account predicts that adding the 5-cent loss will increase the affective evaluability of the $9 loss and therefore be less attractive, whereas the evoked reference set account predicts that the 5-cent loss will evoke an inferior reference set and thus make the gamble *more* attractive. Following up on equivocal findings in the choice task, Experiment 3 compares attractiveness ratings (which require an implicit reference set) and WTP judgments (which do not) in an otherwise closely matched task. Next, Experiment 4 more directly probes effects of the target gamble on scale interpretation, by asking participants to provide an example of one gamble that they would give a very low rating, and one that they would give a very high rating. Finally, Experiment 5 tests whether the reference sets that different gambles evoke are reasonable. We do so by reversing the process: Rather than manipulate the gambles and infer which reference set is evoked, we manipulate the reference set – whether a set of gambles involves primarily gains or a mix of gains and losses – and see if this influences whether a gamble outcome of $0 is described as "win nothing" or "lose nothing". The results from the experiments provide strong support for the proposed account and indicate that evoked reference sets have large and consistent effects on scale interpretation, but that their effect on the internal representation of value is much less robust.

2. Experiment 1

Our first experiment involves three conditions that have not been tested together previously: The standard gamble (henceforth W9/W0), the small-loss gamble (W9/L0.05), and the "lose nothing" gamble (W9/L0). The first two conditions have been compared (Bateman et al., 2007; Slovic et al., 2002), as have the first and third (Bateman et al., 2007), but the three conditions have not been examined within a single study. The evoked reference set account makes a clear prediction about the ordering of attractiveness ratings in the three conditions: W9/W0 < W9/L0.05 < W9/L0. Recall that Bateman et al. provided different explanations of the difference in results between the first two conditions (affective evaluability) and between the first and third conditions (tone), whereas the evoked reference set account can parsimoniously explain the differences between all three.

In addition, choice data have not previously been collected for the W9/L0 gamble, and we do so in this experiment. Comparing choices (between each gamble and a sure amount) across these three gambles is expected to shed light on any effect of evoked reference sets on the internal representation of value.

Relatedly, we also asked some participants to provide WTP judgments (i.e., the most they would be willing to pay to play the gamble one time). WTP judgments have not previously been used for these gambles, and they are potentially useful because, unlike choices, they provide a scaled response while, unlike attractiveness ratings, they do not require participants to impute contextual meaning to an underspecified subjective scale.

Finally, previous experiments comparing responses between the W9/W0 and the W9/L0.05 gambles have confounded the gamble with the manner in which the second outcome is communicated to participants (Bateman et al., 2007; Slovic et al., 2002). The W9/L0.05 gamble has always been presented explicitly as:

7/36 to win $9
29/36 to lose 5¢

By contrast, when participants have been presented with the W9/W0 gamble, the 29/36 chance of winning nothing is not placed right below the first outcome, but is instead stated in the general instructions to participants (see Bateman et al.'s Fig. 1). Thus, one reason for the small-loss effect, as mentioned by Bateman et al. (see their Footnote 6), could simply be that affective evaluability for the W9/L0.05 gamble is higher because attention is drawn to the small-loss outcome by placing it next to the $9 outcome. To test this possibility, the current experiment made explicit the implicit "win nothing" outcome in Slovic et al. and Bateman et al. by placing "29/36 chance to win $0.00" immediately below "7/36 chance to win $9.00". Read et al. (2017) found that making an implicit $0 outcome explicit can, under certain circumstances, influence intertemporal choices.

2.1. Method

Participants were 308 UC San Diego undergraduates who received partial course credit (65% female, mean age = 20). They were randomly assigned to one of three conditions: W9/W0, W9/L0.05, or W9/L0. The task was a paper-and-pencil survey that was part of a series of experiments taking less than an hour. After a general introductory page, participants in the W9/W0 condition read the following:

> Imagine that you have the opportunity to play the gamble below one time for real money. The outcome is determined by spinning a wheel of fortune with 36 areas of equal size. Seven of the areas are green, and 29 of the areas are red. If the spinner lands on a green area, you win $9.00. If the spinner lands on a red area, you win $0.00.
> So the gamble is this:
> 7/36 chance to win $9.00
> 29/36 chance to win $0.00

In the W9/L0.05 and the W9/L0 conditions, "win $0.00" was replaced by "lose $0.05" and "lose $0.00", respectively. All participants reported how attractive they found the gamble by circling one number on a scale ranging from 1 (not at all attractive) to 20 (very attractive). On the same page, half of the participants in each condition then reported whether they would prefer to play the gamble once or receive $2.00 for sure, and the other half reported WTP by answering the question, "What is the most you would be willing to pay to play this gamble one time?".

2.2. Results and discussion

2.2.1. Attractiveness ratings

Fig. 2 shows the results for the attractiveness ratings ($N = 308$). As predicted by the evoked reference set account, mean ratings conformed to the order W9/W0 < W9/L0.05 < W9/L0 (*Ms* = 8.1, 10.9, 14.1). A one-way ANOVA revealed an effect of Gamble ($F(2, 305) = 26.2$, $p < .001$), and pairwise contrasts showed that each condition mean was different from the other (*ts* > 3.5, *ps* < 0.001, Cohen's *ds* > 0.49). Because participants found the W9/W0 gamble to be less attractive than the W9/L0.05 gamble, even though the "win $0.00" outcome was explicit, the original effect is not due to the fact that "win $0.00" was implicit. Nor is it due to the fact that there is something special about a $0 outcome that hinders affective evaluability, because W9/L0 was considered *more* attractive than W9/L0.05. Moreover, affective evaluability is presumably held constant for the W9/W0 and W9/L0 gambles, and the latter was judged much more attractive than the former.

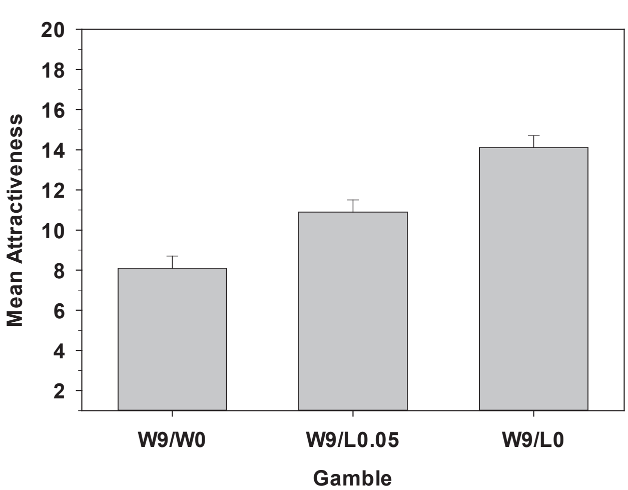

Fig. 2. Experiment 1: Attractiveness ratings as a function of the different gambles. Standard error bars are shown.

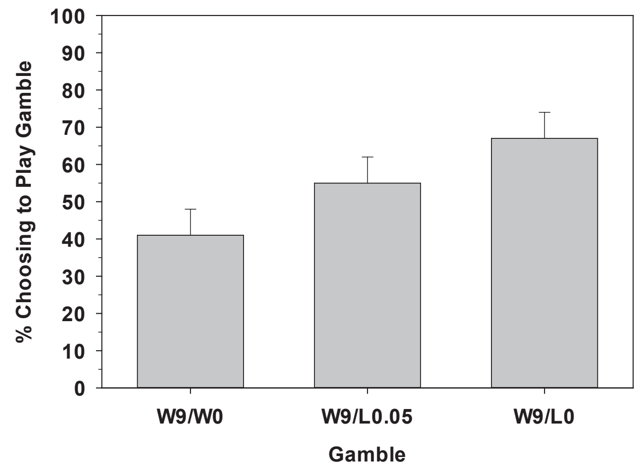

Fig. 3. Experiment 1: Percentage of participants choosing to play each gamble rather than receive $2 for sure. Standard error bars are shown.

2.2.2. Choices

Fig. 3 shows how often participants chose to play each gamble rather than receive $2 for sure ($n = 153$), and the pattern is similar to that for attractiveness ratings. Participants in the W9/W0 condition chose the gamble least often and those in the W9/L0 condition chose it most often (*Ms* = 0.33, 0.45, 0.59). A log-linear analysis showed an effect of Gamble, $\chi^2 = 6.8$, $p = .034$. Separate chi-squared tests, though, revealed that only the W9/W0 and W9/L0 conditions were significantly different from each other, $\chi^2(1, N = 102) = 6.7$, $p = .01$, $\varphi = 0.26$. (For W9/W0 vs. W9/L0.05, $p = .17$, $\varphi = 0.14$, and for W9/L0.05 vs. W9/L0, $p = .22$, $\varphi = 0.12$.) Affective evaluability alone cannot explain the difference between the two $0-outcome gambles. However, the pattern of results is consistent with the evoked reference set account, assuming that reference sets are affecting the internal representation of value.

2.2.3. Willingness to pay

The WTP results were difficult to interpret ($n = 155$). Indeed, 24% of participants reported being willing to pay more than $9 – which is the most they could win by playing the gamble. After eliminating these participants (who were evenly distributed across the three conditions), mean WTP was $2.59, $3.11, and $2.54 for the W9/W0, W9/L0.05, and W9/L0 conditions, respectively. A one-way ANOVA revealed no effect of gamble, $F(2,115) < 1$. One way to interpret this lack of effect is that evoked reference sets affect the construal of ambiguous subjective scales, but that they have little or no effect on the use of a fully interpreted scale like WTP. However, the need to eliminate 24% of responses raises obvious concerns about the meaningfulness of these data. For reasons that are unclear to us, participants apparently had difficulty understanding the WTP question. (We will use a more structured WTP task in Experiment 3.)

In sum, Experiment 1 confirmed the predictions regarding attractiveness ratings and choices made by the evoked reference set account, and cannot be fully explained by affective evaluability. The effects of gamble on choice suggest that part of the effect on attractiveness ratings may be due to inflated internal representation of value, not just scale interpretation. The effects on choice appear less robust than those on ratings, though, which suggests that scale interpretation explains much of the effect on attractiveness ratings. However, the sample size for the choice data was only half that for the ratings data, thereby limiting power.

3. Experiment 2

Experiment 2 was designed to provide an even more direct test of the evoked reference set and affective evaluability accounts. Two new conditions were created that were identical to W9/W0 and W9/L0.05, but instead of a 7/36 chance to win $9, there was a 7/36 chance to *lose* $9. The affective evaluability account predicts that adding the small loss will make the $9 loss more evaluable and therefore attractiveness should *decrease*. By contrast, the evoked reference set account predicts that adding the small loss will evoke a diminished reference set, resulting in *increased* attractiveness ratings for both the $9 win and the $9 loss gambles.

In addition, participants were again asked to choose between one gamble and a sure amount to shed additional light on whether inflated internal representation of value is influencing choices and, by extension, ratings.

3.1. Method

Participants were 220 UCSD undergraduates who received partial course credit (73% female, mean age = 20). They were randomly assigned to one of four conditions (n = 55 in each), and the experiment took the form of a paper-and-pencil survey. In two conditions, participants were presented with either the standard gamble (W9/W0) or the small loss gamble (W9/L0.05) as in Experiment 1. In the other two conditions, the second outcomes were the same as the first two conditions, but the first outcome involved a 7/36 chance to *lose* $9 (L9/W0 and L9/L0.05, respectively). In Experiment 2, as in previous published studies (Bateman et al., 2007; Slovic et al., 2002), the "win nothing" outcome was not explicit (i.e., it was not placed below "7/36 chance to win [lose] $9.00"), but was stated in instructions accompanying the gamble. The stimuli were virtually identical to those presented in Bateman et al.'s Fig. 1 (including a visual depiction of a wheel of fortune). We attempted to replicate earlier published articles as closely as possible. Because the new gambles involved only losses, we changed the attractiveness scale to range from −10 (extremely unattractive) to +10 (extremely attractive). Participants in the two W9 conditions then reported whether they would prefer to play the gamble once or to receive $2 for sure, and those in the two L9 conditions reported whether they would prefer to play the gamble once or to lose $2 for sure.

3.2. Results and discussion

Fig. 4 illustrates the attractiveness results. The left two bars show that the usual effect was replicated: Adding the small loss to the W9/W0 gamble made it more attractive, $t(108) = 3.31$, $p = .001$, $d = 0.63$, $Ms = 0.6, 3.7$. The right two bars represent the critical comparison, and show that adding the small loss to the L9/W0 gamble made it more attractive, as well, $t(107) = 3.55$, $p < .001$, $d = 0.68$, $Ms = -6.5, -2.4$. This is consistent with the evoked reference set prediction and is opposite the affective evaluability prediction.

Fig. 5 shows how often participants preferred to play the gamble (rather than receive $2 for sure in the W9 conditions or lose $2 for sure in the L9 conditions). Adding the small loss increased choices slightly in both the W9 (from 0.44 to 0.49) and the L9 cases (from 0.71 to 0.76), but the increases were not close to significant (χ^2s < 1, φs = 0.05, 0.06). This is inconsistent with affective evaluability, which predicts that the 5-cent loss will increase choices for the W9/W0 gamble and decrease choices for the L9/W0 gamble. These choice results suggest that evoked reference sets did not have much influence on the internal representation of value (broken arrow in Fig. 1). Nonetheless, the results for attractiveness ratings suggest that evoked reference sets exert a consistently strong influence on scale interpretation.

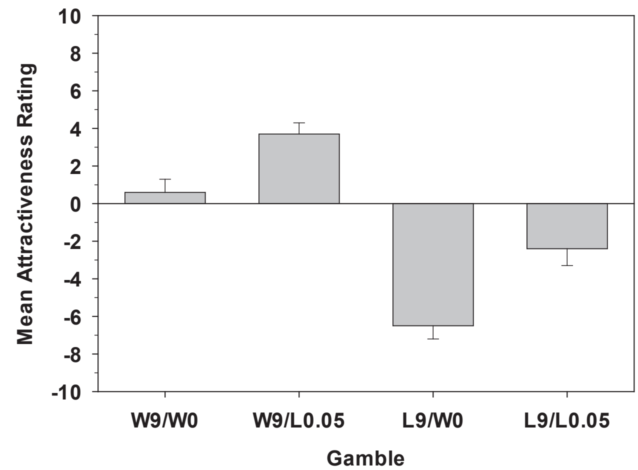

Fig. 4. Experiment 2: Attractiveness ratings as a function of the different gambles. Standard error bars are shown.

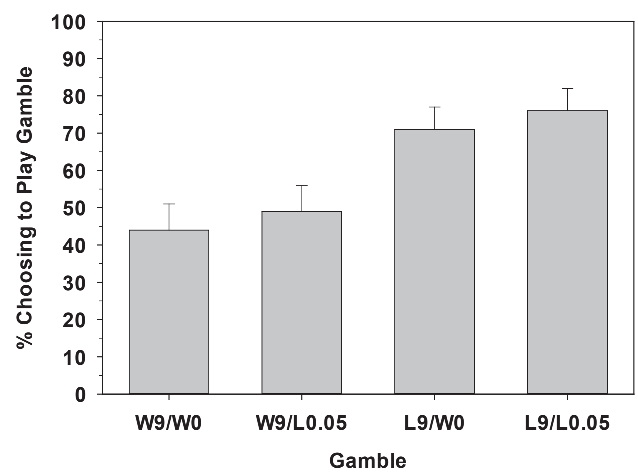

Fig. 5. Experiment 2: Percentage of participants who preferred the gamble to either a sure gain of $2 (left two bars) or a sure loss of $2 (right two bars). Standard error bars are shown.

These results also speak to whether adding the small loss affects how participants interpret the $9 outcome (as predicted by the affective evaluability account) or interpret the gamble more generally (as predicted by the evoked reference set account). On the former account, the effects would have gone in opposite directions for the $9 gain and the $9 loss gambles. The fact that adding a small loss increased ratings for both gambles indicates that the evaluability of the $9 outcome is not central to the effect.

4. Experiment 3

The attractiveness results of Experiments 1 and 2 support the evoked reference set account of why seemingly inconsequential changes to gambles have large effects. A single account can explain both why adding a small loss to a gamble increases its rated attractiveness, as well as why there is a large difference in response to gamble descriptions that are logically equivalent – i.e., the W9/W0 and W9/L0 gambles.

The choice results also support the evoked reference set account over affective evaluability, but they were less clear with respect to whether evoked reference sets were influencing choices. Because no rating scale interpretation is required for the choice task, any influence of reference sets on choice suggests that the effect is driven via the internal representation of value (the broken arrow in Fig. 1). This is not only difficult to justify normatively, but it would also suggest that this path in the model may be influencing attractiveness ratings, over and above the (necessary) effect of scale interpretation. While the effects on attractiveness ratings in Experiments 1 and 2 were large and consistent, the effects on choices were less robust. The results of Experiment 1 seemed to indicate that evoked reference sets were influencing choice (Fig. 3), but only one of the three pairwise comparisons was significant (W9/W0 vs. W9/L0). In Experiment 2, adding the small loss had no significant effect on choices for either the W9/W0 gamble or the L9/W0 gamble, though the small effects were in the direction expected if reference sets were influencing the internal representation of value (Fig. 5).

Comparing results between attractiveness ratings and choices is complicated by the fact that the former are on an interval scale and the latter are dichotomous. In the current experiment, we examined attractiveness ratings and WTP in order to see if adding a small loss affects a well-defined, scaled variable. We changed the WTP directions and question in a manner intended to elicit more meaningful responses than in Experiment 1. Furthermore, both the attractiveness and the WTP response scales were modified in order to make them similar to each other – both were essentially 21-point scales. If we find that adding the small loss continues to have a large effect on attractiveness ratings, but has a small or no effect on WTP, this would provide additional evidence that adding the small loss affects how value is expressed on a subjective response scale, but has little effect on the internal representation of value.

4.1. Method

Participants were 280 UCSD undergraduates (76% female, mean age = 20) who received partial course credit for participation. They were randomly assigned to one of four conditions. Participants in the two attractiveness conditions were presented with either the standard gamble (W9/L0) or the small-loss gamble (W9/L0.05) and rated its attractiveness on a 21-point vertical scale, ranging between 0 at the top (“not at all attractive”) and 21 at the bottom (“extremely attractive”; see Appendix A). Participants placed an X next to one number. The two WTP conditions presented either the standard or small-loss gamble and asked participants to indicate the largest amount of money they would be willing to pay to play the gamble one time. They responded on a 21-point vertical scale that ranged from $0.00 to $10.00 in $0.50 increments by placing an X next to one of the dollar values (see Appendix A).

4.2. Results and discussion

The results are displayed in Fig. 6. The left side shows that the usual finding was replicated using the slightly modified attractiveness scale: Adding the small loss increased attractiveness ratings from 8.5 to 11.4, $t(138) = 2.79$, $p = .006$, $d = 0.47$. The right side shows the results for the WTP conditions, with responses converted to the 21-point scale. Adding the small loss did not significantly increase WTP, $t < 1$, $d = 0.14$, *Ms* = 5.8 ($2.90) and 6.5 ($3.25). It is worth noting that this modified WTP task led to much more coherent responses; only 3.6% of participants (5 out of 140) reported being willing to pay more than $9, compared to 24% in Experiment 1. Eliminating these 5 participants led to almost identical WTP for the standard and small-loss gambles, $2.80 (rating = 5.59) and $2.84 (rating = 5.67), respectively.

In Supplementary Material available online, we report two additional experiments that examined the effect of gamble on choice. Experiment S1 asked 249 participants to choose between receiving $2 for sure and playing either the W9/W0, W9/L0.05, or W9/L0 gamble. Unlike previous experiments, there was no ratings task, only a choice,

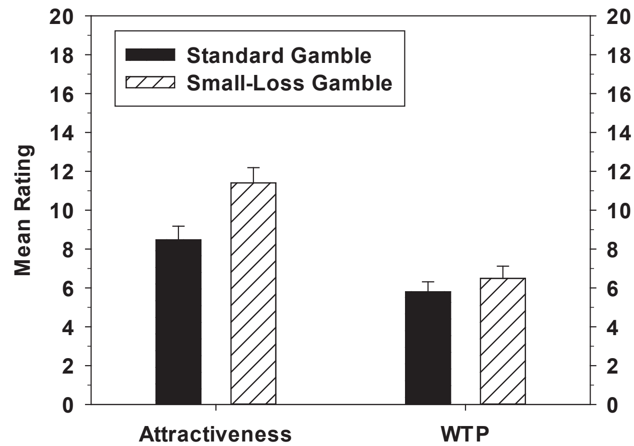

Fig. 6. Experiment 3: Mean attractiveness ratings for the standard and small-loss gambles (left two bars) and mean willingness-to-pay (WTP; right two bars). WTP was converted to the 21-point scale used by participants, where each point corresponded to a $0.5 dollar increment (e.g., a WTP of $3 corresponds to a rating of 6). Standard error bars are shown.

and the results showed no significant effect of the gamble. Because we found the smallest effect on choices when there was no ratings task (Experiment S1) and the largest effect when ratings came first (Experiment 1), it was conceivable that the obligatory use of a reference set in the ratings task was influencing subsequent choices. To test this, we manipulated whether ratings or choice came first in Experiment S2. In this experiment, the choice was between playing one of the three (randomly assigned) gambles and receiving $3 for sure. The results showed that the order of response mode did not matter, and W9/W0 was chosen significantly less often than both W9/L0.05 and W9/L0, which were not significantly different from each other. Taken together, the choice and WTP results from Experiments 1–3 and Experiments S1-S2 indicate inconsistent and relatively small effects of gamble on the internal representation of value, but consistent and large effects on scale interpretation.

5. Experiment 4

Experiment 4 was designed to provide a more direct test of the claim that adding the small loss affects attractiveness ratings because it influences how participants interpret the subjective rating scale. Participants were first asked to rate either the W9/W0 or the W9/L0 gamble on the 0–20 attractiveness scale. They were then asked to provide an example of a gamble to which they would give a very low rating (1 out of 20) and an example of a gamble to which they would give a very high rating (19 out of 20). These examples allow us to compare the reference sets (what counts as a very unattractive and very attractive gamble?) that are induced in the two conditions. We expected that reference sets would be shifted upward in the W9/W0 condition, where the initial gamble is described in “win-win” terms, relative to the W9/L0 condition, where the same gamble is described in “win-lose” terms.

5.1. Method

Participants were 140 UCSD undergraduates who received partial course credit. Eleven participants were excluded because of responses that were either incorrectly formatted (e.g., missing words) or evidently confused (e.g., a high-rated gamble dominated by the low-rated gamble), leaving 129 participants (70% female, mean age = 20) in the analyses reported below. Participants were assigned to one of two conditions, in which they first rated either the W9/W0 gamble (n = 63) or the W9/L0 gamble (n = 66) on the 21-point scale, where 0 = "not all attractive" and 20 = "extremely attractive".

On the next page, they were then asked to provide an example of a gamble that they would rate as 1 on the scale, and to provide an example of a gamble that they would rate as 19. To do so, participants filled in the blank spaces corresponding to two outcomes in a gamble description. The probabilities were provided and were the same as the probabilities in the original gamble they rated. That is, participants wrote in outcomes ("win X" or "lose X") in the blank fields in a gamble description with the form: "A 7/36 chance to [blank field] and a 29/36 chance to [blank field]." (The detailed instructions provided to participants are included in Appendix B.)

5.2. Results and discussion

We first consider attractiveness ratings in the initial task. Replicating the results of Experiment 1, participants again gave higher mean ratings to the W9/L0 gamble (M = 15.2) than to the W9/W0 gamble (M = 9.0), $t(127) = 5.72$, $p < .001$, d = 1.01.

Next, we turn to the examples of gambles that would receive low and high ratings on the attractiveness scale. We classified each gamble (1-rated and 19-rated) according to the sign of the objective values (win, lose, zero) of its two outcomes. (Note that "win 0" and "lose 0" are coded equivalently as objective zero outcomes, because we are interested in whether reference sets contain different gambles in the two conditions, not in whether participants describe the same gambles differently.) For example, if a gamble included one win and one zero outcome (in either order), it would be coded as a "Win/Zero" gamble, while a gamble in which both outcomes involved positive gains would be coded as a "Win/Win" gamble.

The distributions of participants' 1-rated and 19-rated gambles are summarized in Tables 1 (for the W9/L0 condition) and 2 (for the W9/W0 condition). Note that cells farther to the right and lower down in the tables correspond to better reference sets. To ask whether losses are more likely to be included in reference sets for the W9/L0 gamble, we focus first on the proportion of participants for whom at least one gamble included at least one objective loss component (i.e., Win/Lose, Lose/Zero, or Lose/Lose). In the W9/L0 condition, 86.4% (57/66) of participants provided at least one gamble with a possible loss, while in the W9/W0 condition, only 49.2% (31/63) of participants provided at least one gamble with a possible loss ($p < .0001$, two-tailed Fisher's exact test, $\varphi = 0.40$). That is, the use of a loss frame in the focal gamble tends to draw losses into the reference set that defines the attractiveness scale. A second noteworthy contrast between the distributions shown in Tables 1 and 2 concerns the frequency of Win/Win gambles, in which both outcomes are non-zero gains. In the W9/L0 condition, only 15.2% (10/66) of participants generated at least one Win/Win example, while 55.6% (35/63) of participants in the W9/W0 condition did so ($p < .0001$, $\varphi = 0.42$).

The framing of the initial gamble – where the second outcome is either "win \$0" or "lose \$0" – thus has a clear effect on interpretations of the rating scale (what counts as a low-rated or a high-rated gamble on the scale?). After seeing the W9/L0 gamble, participants are far more likely to assume that Win/Lose gambles are included in the reference set, and far less likely to assume that Win/Win gambles are. The different reference sets induced by the two initial gambles, in turn, explain the enhanced attractiveness ratings observed in the "lose \$0" frame.

6. Experiment 5

Thus far we have tested the competing accounts in part by manipulating gamble descriptions and seeing how this affects which reference sets are apparently evoked. In Experiment 5, we reversed the process: We manipulated a gamble's reference set to see how this affects the gamble's description. The evoked reference set account assumes that gamble descriptions evoke certain reference sets because the descriptions are a signal from a speaker (the experimenter, in this case) about the reference set. For example, it assumes that speakers are more likely to describe a \$0 outcome as "win \$0" (rather than "lose \$0") when the gambles in the reference set consist of wins, compared to when the gambles in the set consist of both wins and losses. We noted earlier that the objective components of a gamble provide information about the reference set from which it was selected (e.g., a gamble involving wins and losses could not have been sampled from a set of gambles involving only wins). Adapting a frame selection task used in earlier research on "information leakage" (Sher & McKenzie, 2006), the current experiment tests the prediction that how a gamble is framed also provides information about the relevant reference set. If the prediction is confirmed, it would indicate that the reference sets that different gamble descriptions evoke are not only predictable, but reasonable.

6.1. Method

Participants were 110 UCSD undergraduate students who received partial course credit (66% female, mean age = 20). They read over a table of 12 gambles with the knowledge that one of the gambles would be selected and they would have to describe that gamble to a friend, who would then decide whether to play the gamble. (For details, see Appendix C.) The gambles in the table were composed of two possible outcomes and their probabilities (e.g., Gamble 10 was a 7/36 chance of \$9 and a 29/36 chance of \$0). The first outcome was always a gain (i.e., a positive dollar amount). Participants randomly assigned to the Win/Win condition (n = 55) saw gambles whose second outcome was almost always a gain (only 2 of the 12 values were negative), whereas those in

Table 1
Frequencies of 1- and 19-rated gamble examples in the W9/L0 condition (n = 66).

		19-Rated Gamble					
		Lose/Lose	Lose/Zero	Win/Lose	Zero/Zero	Win/Zero	Win/Win
1-Rated Gamble	Lose/Lose	0	0	**3 (4.5%)**	0	0	**2 (3.0%)**
	Lose/Zero	0	0	**6 (9.1%)**	0	**5 (7.6%)**	0
	Win/Lose	0	0	**15 (22.7%)**	0	**19 (28.8%)**	**6 (9.1%)**
	Zero/Zero	0	0	0	0	**1 (1.5%)**	0
	Win/Zero	0	0	**1 (1.5%)**	0	**6 (9.1%)**	**2 (3.0%)**
	Win/Win	0	0	0	0	0	0

Table 2
Frequencies of 1- and 19-rated gamble examples in the W9/W0 condition (n = 63).

		19-Rated Gamble					
		Lose/Lose	Lose/Zero	Win/Lose	Zero/Zero	Win/Zero	Win/Win
1-Rated Gamble	Lose/Lose	0	0	0	0	0	**1 (1.6%)**
	Lose/Zero	0	0	0	0	**1 (1.6%)**	0
	Win/Lose	0	0	**5 (7.9%)**	0	**11 (17.5%)**	**9 (14.3%)**
	Zero/Zero	0	0	0	0	0	0
	Win/Zero	0	0	**2 (3.2%)**	0	**9 (14.3%)**	**21 (33.3%)**
	Win/Win	0	0	**2 (3.2%)**	0	0	**2 (3.2%)**

the Win/Lose condition (n = 55) saw gambles whose second outcome was never a gain. In the latter case, 11 of the 12 second outcomes were negative, and one (Gamble 10) was $0. The two tables were identical except that all the positive second outcomes in the Win/Win table were made negative by placing a negative (−) sign in front of each to create the Win/Lose table. Gamble 10, however, was identical in both tables.

After reading the table, participants were asked to describe Gamble 10 to a friend (see Appendix C). They had to fill in the probability, dollar value, and circle the valence (win vs. lose) for both outcomes. For half of the participants, "win" was on top and "lose" was on the bottom; this was reversed for the other half. The key dependent measure was how often participants described the second outcome as a chance to "win" $0 vs. "lose" $0.

6.2. Results and discussion

Three participants were excluded from the analysis because they did not fully complete the questionnaire. In the Win/Win condition, 64% of participants described the $0 outcome as "win $0", whereas in the Win/Lose condition, only 24% did so, $\chi^2(1, N = 107) = 17.44$, $p < .001$, $\varphi = 0.40$. As expected, in a context primarily involving gains, the $0 outcome was more likely to be described as winning nothing compared to a context involving gains and losses. This indicates that describing the second outcome as "win nothing" – as the standard gamble does – should signal to the listener that the relevant reference set of gambles primarily involves gains (rather than gains and losses). Similarly, describing the outcome as "lose nothing" should signal that the reference set of gambles involves a mixture of gains and losses (rather than only gains).

7. General Discussion

The present studies build on an intriguing finding reported by Slovic et al. (2002), in which adding a small loss to a gamble increases its judged attractiveness. Since the small-loss gamble is strictly worse than the standard gamble, these between-subjects results are puzzling. Slovic et al. and Bateman et al. (2007) appealed to the affect heuristic to explain the finding, with the idea being that adding the small loss makes the $9 outcome affectively evaluable. Bateman et al. also presented a gamble to participants with the second outcome described as "lose nothing" rather than "win nothing", and ratings were even higher, even though affective evaluability is presumably held constant for these two $0-outcome gambles. Bateman et al. instead suggested that the more positive tone of "lose nothing" compared to "win nothing" explained the difference.

We proposed that the pattern of results can be explained more parsimoniously by positing that different outcome valences (win vs. lose) evoke different reference sets for comparison (Kahneman & Miller, 1986; Leong et al., 2019; McGraw et al., 2010; see also Biernat et al., 1991). When evaluating a single gamble in isolation on the subjective attractiveness scale, participants must answer the implicit question, "How attractive compared to what?". Because the standard gamble (W9/W0) involves only wins, participants are more likely to compare it to other gambles involving only wins. However, adding a small loss (W9/L0.05) or reframing the neutral outcome as a zero loss (W9/L0) results in a different reference set for comparison, namely, gambles involving wins and losses. The downward-shifted reference set in the two loss conditions elevates the rated attractiveness of any fixed gamble, resulting in a small loss advantage. The evoked reference set account predicts a specific order of attractiveness judgments (W9/W0 < W9/L0.05 < W9/L0) that was confirmed in Experiment 1.

Experiment 2 added two new conditions to examine the effect of adding a small loss to the risk of a $9 *loss*. The affective evaluability account predicts that the small loss should make the $9 loss more evaluable and therefore make the gamble less attractive. By contrast, the evoked reference set account predicts that, because the addition of the small loss shifts the reference set from win-lose to lose-lose gambles, it should result in an increase in attractiveness ratings. The results clearly supported the evoked reference set account.

The current evoked reference set account is consistent with Kahneman and Miller's (1986) norm theory. They speculated that the standard gamble and the small-loss gamble might evoke different norms (analogous to reference sets), and our article is the first to report evidence for the account and distinguish it from the competing affective evaluability account. As we discuss below, we have also extended the norm theory account by (1) showing that the standard and small-loss gambles do in fact evoke different reference sets when providing attractiveness ratings, (2) showing that the reference sets evoked by different gamble descriptions are well-attuned to the effects of reference sets on speakers' gamble descriptions, and (3) distinguishing between psychologically and normatively distinct pathways – scale interpretation and internal representation of value – through which reference sets may affect evaluation, and we assess their contributions by comparing ratings with choices and WTP.

Not only does the evoked reference set account naturally explain the full set of attractiveness ratings, it also casts doubt on the claim that the effect is counter-normative. Interpreting the subjective, bounded attractiveness scale requires implicitly, if not explicitly, answering the question of what the range of relevant gambles might be. With only a single gamble to evaluate, participants can only answer the question in rough and uncertain terms. They appear to make a parsimonious inference: If the lone gamble includes references to both wins and losses, participants are far more likely to assume that win-lose gambles are included in the reference set, and far less likely to assume that win-win gambles are (Experiment 4). This inference from gamble frame to underlying reference set seems to be ecologically appropriate. We showed in Experiment 5 that these logically equivalent gamble frames leak reference set information: Speakers are more likely to describe a $0 outcome as "win $0" (rather than "lose $0") when the other gambles in a reference set involve primarily wins as opposed to when the other gambles in the set involve a mixture of wins and losses.

However, there are two ways in which evoked reference sets can influence subjective ratings – via the interpretation of the scale, or via the internal representation of the gamble's value (Fig. 1). While the scale interpretation route is both necessary and normatively unproblematic, effects on perceived value are psychologically plausible but not compatible with standard normative models of choice. We attempted to determine the extent to which each route is affecting ratings by studying choice and WTP, which eliminate the rating scale interpretation route. We found that the effects of the gamble on attractiveness ratings were robust and consistent, while the effects on choices were relatively weak and often statistically non-significant in Experiments 1 and 2, and there was no significant effect on WTP in Experiment 3. Nonetheless, even when effects on choices (and WTP in Experiment 3) were not significant, they were always in the direction predicted by evoked reference sets. Thus, our data suggest that evoked reference sets have some influence on choice, but the effect may be relatively small and inconsistent. These results are interesting not just with respect to choices per se, but also their implications for attractiveness ratings. Because the choice task eliminates rating scale interpretation, the effects on choice indicate that evoked reference sets influence the internal representation of value. However, in the ratings task, both scale interpretation and inflated internal representation of value are potential routes to a small loss effect, and the effects on ratings are large and consistent, suggesting that much of the work is being done by scale interpretation.

If the choice effect is real, it suggests a subjective contrast effect, whereby inferior evoked reference sets enhance the perceived value of the target gamble (broken arrow in Fig. 1). Though we do not posit a specific mechanism for such a subjective contrast effect, it is compatible with several models of evaluation and choice in the literature. According to range-frequency theory, an inferior reference set for the evaluation of gamble outcomes would improve the range-normalized position as well as the percentile of both outcomes, leading to enhanced perception of the the value of the gamble as a whole (Mellers & Cooke, 1994). Alternatively, effects on the internal representation of value could be explained via decision by sampling (DbS) theory, according to which the gamble's gains are ordinally compared to a sample of other possible gains stored in memory, while losses are similarly compared to remembered losses (Stewart et al., 2006; Walasek & Stewart, 2015, 2019). The more often an outcome comes out ahead in these binary comparisons, the greater its subjective value. Assuming that the "lose" ("win") frame invites comparison of the small outcome to other losses (wins), small-loss gambles should be subjectively superior. We note, however, that while the DbS mechanism could generate effects on choice (which we sometimes found) and WTP (which we did not find), it seems less relevant to the target gamble's effect on interpretations of low and high rating scale levels in Experiment 4. Importantly, this sizeable effect on scale interpretation suffices to explain the effects on attractiveness ratings that we consistently found.

The model in Fig. 1 distinguishes between a gamble's objective content (probabilities and outcomes) and how it is described (e.g., "win 0" vs. "lose 0"). Both components may convey relevant information about the underlying reference set. Information conveyed by the objective content is consistent with the "options as information" model (Sher & McKenzie, 2014), which provides a rational analysis of some apparently counter-normative context effects in human decision making. Researchers often compare decisions and evaluations across different choice sets, and regard inconsistent ordering across contexts as evidence for irrationality. However, when – as is often the case in these studies – the natural space of options is poorly known, participants may reasonably treat the choice set as a sample from this space. Different choice sets may then lead to different inferences, which may in turn lead to different preferences. For example, joint-separate reversals occur when option A is rated higher than B when each is evaluated separately (i.e., between-subjects), but B is rated higher than A when evaluated jointly (i.e., within-subjects; e.g. Hsee, 1996). However, Sher and McKenzie (2014) showed that these findings are accounted for by a rational model in which judges learn from the presented options (about the distribution of attribute values) and update their preferences accordingly: Participants drew markedly different inferences from different (separate and joint) evaluation sets, and, when these inferences were presented as background information to a different group of participants, they sufficed to reproduce the joint-separate effect. A similar analysis has been applied to apparently intransitive choice behavior, in which choices made in multiple pairwise contexts are not compatible with a single underlying preference order (Müller-Trede, Sher, & McKenzie, 2015), and the phenomenon of asymmetric dominance, in which the addition of a third inferior "decoy" option to a "core" two-option choice set systematically alters preferences over the core options (Kamenica, 2008; Prelec, Wernerfelt, & Zettlemeyer, 1997; Ratneshwar, Shocker, & Steward, 1987; Sher, Müller-Trede, & McKenzie, Unpublished manuscript). In these tasks, as in the gamble evaluation task, participants draw different – and reasonable – inferences depending on the specific options presented to them, and their expressed preferences and attitudes reflect this.

The influence of the gamble's *description* in Fig. 1, on the other hand, is an example of "information leakage", a normative framework we have developed to explain some framing effects (Sher & McKenzie, 2006, 2008, 2011; see also Keren, 2007; McKenzie, 2004; McKenzie, Liersch, & Finkelstein, 2006; McKenzie & Nelson, 2003; Teigen & Karevold, 2005). The approach uses conversational pragmatics to shed light on why logically equivalent utterances result in different listener behavior (Grice, 1975; Hilton, 1995; Schwarz, 1994). Generally, the fact that a speaker (e.g., experimenter) chooses to describe an option or outcome in a particular way (e.g., "win 0" vs. "lose 0", or beef that is "85% lean" vs. "15% fat") can convey relevant information to listeners. For example, a new medical treatment that leads to more survivors than other treatments is relatively likely to be described as leading to a "50% survival rate" (rather than a "50% mortality rate"). That is, the distribution of other treatments' efficacy influences a speaker's choice of description. Furthermore, listeners are sensitive to the speaker's choice of description. When a new treatment is described as having a "50% survival rate", listeners are relatively likely to infer that other treatments lead to lower survival rates. That is, describing a treatment in terms of its "survival rate" triggers an inference to a reference set of less effective treatments.

When seemingly inconsequential changes to the decision environment affect judgments and decisions, the result is usually construed as evidence of irrationality. However, it is necessary to rule out the possibility that the change in context provides information that a rational actor would utilize (Sher & McKenzie, 2008, 2011), and it is well known that decision makers can be sensitive to even subtle changes in context (e.g., Payne et al., 1993). In this spirit, the present model distinguishes between normative and non-normative ways in which adding a small loss to a gamble can increase its attractiveness when evaluated in isolation. Data from our five experiments suggest that the higher attractiveness ratings for the dominated, small-loss gamble make sense, in part, because participants are making parsimonious inferences about the range of outcomes given the single gamble presented to them. This, in turn, influences the interpretation of the subjective rating scale in reasonable ways. Data from choices, however, indicate that the evoked context can also influence the internal representation of value, though this non-normative effect is much less robust in the present task. Adding a possible small loss does not just make a gamble slightly worse. It also conveys information about the task-relevant comparison set, relative to which the same gamble may emerge as far better.

8. Author note

This research was supported by National Science Foundation Grants SES-0820553, SES-1060270, and a Scholar Award (to the second author) from the James S. McDonnell Foundation.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Appendix A. Experiment 3 stimuli

After brief directions, participants in the two attractiveness ratings conditions answered the question below. Participants saw the second component of the gamble with either “win $0.00” or “lose $0.05”:

Imagine you have the opportunity to play this gamble one time:

7/36 chance to win $9.00
29/36 chance to win $0.00 [lose $0.05]

How attractive do you find the gamble? In the space below, please place an X next to one number, where 0 = “not at all attractive” and 20 = “extremely attractive”.

0	_____	*not at all attractive*
1	_____	
2	_____	
3	_____	
4	_____	
5	_____	
6	_____	
7	_____	
8	_____	
9	_____	
10	_____	
11	_____	
12	_____	
13	_____	
14	_____	
15	_____	
16	_____	
17	_____	
18	_____	
19	_____	
20	_____	*extremely attractive*

Participants in the two WTP conditions answered the question below. Participants saw the second component of the gamble with either "win $0.00" or "lose $0.05":

Imagine you could buy a ticket that allows you to play this gamble one time:

7/36 chance to win $9.00
29/36 chance to win $0.00 [lose $0.05]

What is the most you would be willing to pay to buy a ticket for the gamble? In the space below, please place an X next to the largest price you would be willing to pay for a ticket.

By placing an X next to a price, you are indicating that you would be willing to pay that price for the ticket, but that you would not be willing to pay any larger price for the ticket.

$0.00 _____
$0.50 _____
$1.00 _____
$1.50 _____
$2.00 _____
$2.50 _____
$3.00 _____
$3.50 _____
$4.00 _____
$4.50 _____
$5.00 _____
$5.50 _____
$6.00 _____
$6.50 _____
$7.00 _____
$7.50 _____
$8.00 _____
$8.50 _____
$9.00 _____
$9.50 _____
$10.00 _____

Appendix B. Experiment 4 stimuli

After participants rated the attractiveness of either the W9/W0 standard gamble or the W9/L0 gamble, they were presented with the following (with "win $0.00" replaced with "lose $0.00" in the W9/L0 condition):

On the previous page, you used a 0-20 scale to rate the attractiveness of a gamble. You were given a gamble and asked to rate it. Now, we will give you a rating, and ask you to provide an example of a gamble that, on the same 0-20 scale, would get that rating.

Here is the gamble you just rated:

A 7/36 chance to [win $9.00] and a 29/36 chance to [win $0.00]

We have highlighted the two possible outcomes of the gamble. In the questions below, please write in outcomes to create a gamble that you would give the indicated rating to.

For example, in the first question below, you are asked to provide an example of a gamble that you would rate as a "1". To keep things simple, we have already written out the chances for the outcomes. You just need to fill in the two outcome values, to create an example of an overall gamble that you would rate as a "1" on the 0-20 scale you saw earlier.

The second question is similar, except that you are asked to provide an example of a gamble that you would rate as a "19" on the same 0-20 scale.

If you have any questions, please ask the experimenter.

Example of a gamble you would rate as "1":

A 7/36 chance to [ ] and a 29/36 chance to [ ]

Example of a gamble you would rate as "19":

A 7/36 chance to [ ] and a 29/36 chance to [ ]

Appendix C. Experiment 5 stimuli

After a short page of general instructions, participants read the following on page 2:

In this experiment, you will read about a number of possible gambles. Then one of the gambles will be selected. You will be asked to describe this selected gamble to someone else.

The Table on the next page lists all of the possible gambles. For each gamble, there are two potential *outcomes*. Each outcome is some amount of money the person playing the gamble could win or lose. If the amount of money is positive, this means the person wins that amount of money. If the amount is negative, this means the person loses that amount of money.

In every gamble, each of the two monetary outcomes has a certain *probability*. This probability indicates how likely it is that a person playing this gamble will have this monetary outcome.

For example, please look at **Gamble 4** in the Table on the next page. In this gamble, a person will win $18 with probability 1/4, and will lose $5 with probability 3/4.

In other words, 1 out of every 4 times a person plays Gamble 4, they will win $18, and the rest of the time they play Gamble 4 they will lose $5.

Now please turn the page, and carefully read through all of the 12 gambles listed in the Table.

If you don't understand anything about the Table, feel free to turn back to this page to re-read the explanation above.

On the third page was the table listing 12 gambles and further instructions. Below is the table presented to the "Win/Win" participants. The table presented to the "Win/Lose" participants was identical except that all of the non-zero values in the Outcome 2 column were negative (i.e., were losses).

Table of Gambles

Gamble	**Probability of Outcome 1**	**Outcome 1**	**Probability of Outcome 2**	**Outcome 2**
1	2/5	$10	3/5	$2
2	7/14	$5	7/14	$12
3	2/3	$20	1/3	$3
4	1/4	$18	3/4	–$5
5	19/40	$8	21/40	$11
6	90/100	$0	10/100	$17
7	14/16	$3	2/16	–$8
8	3/8	$4	5/8	$1
9	1/12	$3	11/12	$4
10	7/36	$9	29/36	$0
11	5/9	$4	4/9	$3
12	3/20	$16	17/20	$2

When you have finished reading all the gambles in the Table, please turn to the next page.

On the next page, participants were presented with the following (though for half the participants the options to be circled had "lose" on top and "win" on the bottom):

Now imagine you had to describe **Gamble 10** to a friend, who will then have to decide whether to play this gamble.

Please review Gamble 10 in the Table on the previous page, and then complete the following description as appropriate.

"In Gamble 10, there is a _____/_____ *(write #'s)* chance that you will win / lose *(circle one)* $_________ *(write #)*,

and there is a _____/_____ *(write #'s)* chance that you will win / lose *(circle one)* $_________ *(write #)*."

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104043.

References

Alempaki, D., Canic, C., Mullett, T. L., Skylark, W. J., Starmer, S., Stewart, N., & Tufano, F. (2019). Re-examining how utility and weighting functions get their shapes: A quasi-adversarial collaboration providing a new interpretation. *Management Science* (in press).

Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making, 20*, 365–380.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgments. *Journal of Personality and Social Psychology, 60*, 485–499.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*, 243–249.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.). *Syntax and semantics volume 3: Speech acts*. New York: Academic Press.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*, 247–257.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*, 90–98.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Kamenica, E. (2008). Contextual inference in markets: On the informational content of product lines. *American Economic Review, 98*, 2127–2149.

Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes, 103*, 238–255.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making, 30*, 1147–1156.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2019). Illusory inconsistencies in judgment: Stimulus-evoked reference sets and between-subjects designs. *Psychonomic Bulletin and Review, 26*, 647–653.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.

McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science, 21*, 1438–1445.

McKenzie, C. R. M. (2004). Framing effects in inference tasks – and why they are normatively defensible. *Memory & Cognition, 32*, 874–885.

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review, 10*, 596–602.

McKenzie, C. R. M., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics, 5*, 337–360.

Mellers, B. A., & Cooke, A. D. J. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1055–1067.

Müller-Trede, J., Sher, S., & McKenzie, C. R. M. (2015). Transitivity in context: A rational analysis of intransitive choice and context-sensitive preference. *Decision, 2*, 280–305.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*(6), 407–418.

Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications.* Mahwah, NJ: Erlbaum.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge: Cambridge University Press.

Prelec, D., Wernerfelt, B., & Zettlemeyer, F. (1997). The role of inference in context effects: Inferring what you want from what is available. *Journal of Consumer Research, 24*, 118–125.

Ratneshwar, S., Shocker, A. D., & Steward, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research, 13*, 520–533.

Read, D., Olivola, C. Y., & Hardisty, D. J. (2017). The value of nothing: Asymmetric attention to opportunity costs drives intertemporal decision making. *Management Science, 63*, 4277–4297.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology, 26*, 123–162.

Sher, S., Müller-Trede, J., & McKenzie, C. R. M. (2016). The rational construction of preference. Unpublished manuscript.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494.

Sher, S., & McKenzie, C. R. M. (2008). Framing effects and rationality. In M. Oaksford, & N. Chater (Eds.). *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 79–96). Oxford: Oxford University Press.

Sher, S., & McKenzie, C. R. M. (2011). Levels of information: A framing hierarchy. In G. Keren (Ed.). *Perspectives on framing* (pp. 35–64). Psychology Press - Taylor & Francis Group.

Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General, 143*, 1127–1143.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). Cambridge: Cambridge University Press.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1–26.

Stewart, N., Reimers, S., & Harris, A. J. L. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61*, 687–705.

Teigen, K. H., & Karevold, K. I. (2005). Looking back versus looking ahead: Framing of time and work at different stages of a project. *Journal of Behavioral Decision Making, 18*, 229–246.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38*, 599–637.

Walasek, L., & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General, 144*, 7–11.

Walasek, L., & Stewart, N. (2019). Context-dependent sensitivity to losses: Range and skew manipulations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 957–968.

Wernerfelt, B. (1995). A rational reconstruction of the compromise effect: Using market data to infer utilities. *Journal of Consumer Research, 21*, 627–633.

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin, 139*, 497–518.