Evaluations Are Inherently Comparative, But Are Compared To What?

Minah H. Jung, ¹ Clayton R. Critcher², Leif D. Nelson²

¹New York University, ²University of California, Berkeley

Authors’ Note

Minah H. Jung, Leonard N. Stern School of Business, New York University; Clayton R. Critcher, Haas School of Business, University of California, Berkeley; Leif D. Nelson, Haas School of Business, University of California, Berkeley.

Correspondence concerning this article should be addressed to Minah H. Jung, Leonard N. Stern School of Business, New York University, 40 West 4th Street, Tisch Hall 912, New York, NY 10012. E-mail: minah.jung@stern.nyu.edu.
Abstract

Understanding how objective quantities are subjectively characterized has been a central topic of investigation in psychology. Decision by sampling has offered the first comprehensive account for how objective stimuli are subjectively evaluated to guide decisions. That theory suggests an inherently comparative procedure: Values seem larger or smaller based on how they rank in a comparative set, the decision sample. Although decision by sampling has proven its practical utility in several ways, its application is limited by offering an incomplete answer to a central question: What values are included in the decision sample? We identify and test four accounts, each suggesting that how values are processed determines whether they linger in the sample. Testing our ideas through studies of loss aversion and temporal discounting, we find clear support for one account: Quantities need to be subjectively evaluated—rather than merely seen—for them to enter the sample and guide decision making.
Evaluations Are Inherently Comparative, But Are Compared To What?

In order to make good decisions, people need to evaluate the available options. In some cases, choices become easy-to-define (even if difficult-to-calculate) exercises in optimization. For example, bees—like computers—seem particularly adept at solving some problems of this sort: They can evaluate routes when searching for food in order to maximize the calories found minus the calories expended (Real, 1991, 1996).

Other choices—particularly those that involve tradeoffs—are more complex. In deciding whether to upgrade a smartphone or an airline seat, one must consider whether such enhancements are “worth it.” This requires people to characterize objective differences in subjective terms. Understanding how this is done has been a major project for psychologists, economists, and others interested in judgment and decision making (Bernoulli, 1954; Edwards, 1954; Kable & Glimcher, 2007; Kahneman & Tversky, 1979; Parducci, 1968; Stewart, Chater, & Brown, 2006).

Prospect theory emphasized—among other innovations—that valuation is inherently comparative (Kahneman & Tversky, 1979). Two airlines currently offering equivalent legroom may find themselves with differently satisfied customers if one airline arrived at that outcome by adding an inch whereas the other subtracted an inch from a previous layout. That is, identical current values may be evaluated differently when they are compared to different reference points. Numerous framing problems demonstrate how even arbitrary reference points affect valuation (Kühberger, 1998; Levin, Schneider, & Gaeth, 1998; Tversky & Kahneman, 1981).

More recently, decision by sampling (DbS) has been proposed to suggest how comparison processes are more deeply entrenched in valuation (Stewart et al., 2006; see Noguchi & Stewart, 2018, for a critical extension of DbS) than previous theories anticipated. The theory
stands in contrast to a “value-first” approach to decision making (Vlaev, Chater, Stewart, & Brown, 2011)—one in which there is a direct mapping from a specific attribute value to a subjective valuation in the mind. With decision by sampling, attribute values feed into comparison-based decision making (Vlaev et al., 2011), such that the value informs judgments and decisions based on how the value ranks in a set of relevant comparison values. For example, an airline passenger who looks down at her 15 inches of legroom may evaluate the airline’s generosity by calling to mind a sample of previously encountered values against which this new value is compared. If on recent flights she experienced legrooms of 12, 13, 13, 13, and 16 inches, her current seating may seem spacious—better than 4/5 of those in the decision sample. But were she to be returning to the skies after a twenty-year hiatus, her decision sample may be 14, 18, 18, 19, and 19 inches. In this case, her seat may seem cramped, better than only 1/5 of those sampled.

Empirical support for DbS has been varied and substantial. Numerous investigations have documented how judgments and decisions are seemingly informed by comparisons with values to which they are recently or chronically exposed (Olivola & Sagara, 2009; Stewart et al., 2006; Walasek & Stewart, 2015). Although there is a substantial understanding of how decision samples guide evaluation and decision making, less is known about which quantitative values enter such samples. Given the relevance of DbS to decision making in many social and economic domains, addressing this question has broad relevance.

Stewart et al. (2006) say recent exposure likely matters, but “similarity and background knowledge will surely play a role as well” (p. 4). Sometimes, this refers to categorical similarity: Evaluation of salaries and car prices may rely on comparisons with other salaries and car prices, respectively, but not with each other (Rablen, 2008; see also Hounkpatin, Wood, & Dunn, 2016).
It can also mean value similarity: A person considering a $649 mobile phone is more likely to use a $689 phone as a relevant comparison than a $999 phone (Brown & Stewart, 2005; Qian & Brown, 2005). Other research suggests value distinctiveness plays a role: If one is exposed to many highly similar values, the chance that any one of those values enters the decision sample may be reduced (Brown & Matthews, 2011; Brown, Neath, & Chater, 2007; Tripp & Brown, 2016).

What unified these previous efforts is their focus on trying to identify properties of values that predict their inclusion in a decision sample. This approach is limited because outside of experimental contexts people are exposed to such a vast array of quantities it becomes difficult to predict the resulting sample. We instead ask how the processing of values influences their inclusion in a decision sample. Consider a recent paradigm used by Walasek and Stewart (2015), who argued that decision by sampling can help to explain loss aversion. In their research, participants were shown a series of lotteries and indicated which ones that they would play. The authors varied the distribution of losses and/or gains that defined such lotteries. Participants’ loss aversion—their greater sensitivity to changes in losses than changes in gains—changed just as decision by sampling would predict. But why? Participants were exposed to the lotteries, which included values to which they had to be responsive, requiring them to evaluate their attractiveness, and then ultimately make a decision about whether to take the risk. But precisely which one or more of these four steps accounts for the values’ placement in the decision sample, thereby influencing decisions, is unclear. Answering this question should enable a more precise understanding of which values guide comparative decision making.

We conducted experiments to determine whether one or more of these steps predicted what attribute values enter the decision sample. The decision sample is not directly measurable
but can be inferred indirectly. The first two studies disentangle the accounts by using a lottery paradigm modified from that developed by Walasek and Stewart (2015). With those insights in mind, in Study 3, we ported our theory to a novel domain (patience) to see if the insights apply to decision making more generally.

**Study 1**

Studies 1 and 2 lean on Walasek and Stewart’s (2015) paradigm to test what psychological processes are responsible for placing values in decision samples. In their paradigm, participants confront mixed gambles, indicate which ones they would accept, and thereby reveal a loss aversion coefficient. Supplemental Study A offered a more precise test that gave us confidence that participants’ loss aversion varies just as DbS would expect. Supplemental Study B showed that merely being exposed to values was insufficient to place them in the decision sample. Based on this finding, all participants in Study 1 were exposed to the same gain and loss values, but we varied how participants were led to process some of these values. This permitted us to test whether merely responding to the values, subjectively evaluating the values, or making decisions based on the values is what places them in the decision sample.

**Method**

**Preregistration.** For all three studies, we preregistered our sample size, design, and analysis plan. These preregistrations along with full data and materials can be found here: [https://osf.io/cdxsw/?view_only=1890adad9288485eaebc1f953163992b.](https://osf.io/cdxsw/?view_only=1890adad9288485eaebc1f953163992b)

**Participants and design.** Participants were 1,062 Americans recruited from Amazon Mechanical Turk (AMT). They were randomly assigned to one of four gain range conditions:
narrow + exposure, narrow + response, narrow + evaluation, or wide. This sample size was informed by the results of Supplemental Studies A and B, which used a similar paradigm.

**Procedure.** We began by explaining to participants that they would be exposed to a series of lotteries. To reinforce that the chance of winning or losing each lottery was equivalent, we explained that each lottery’s outcome would be determined by the flip of a coin. Heads would produce a win; tails, a loss. Participants saw one example lottery, which illustrated the format that would be used on the target trials.

On every trial, participants were exposed to the same two number lines—one for gains, one for losses. The gain and loss values for a particular lottery were identified on their respective number line. The gain number line spanned from $6 to $32, whereas the loss number line ranged from -$20 to -$6 (see Figure 1). This assured that participants were exposed to the same set of numbers. Note that 112 unique lotteries—defined by one of 14 different gain values and one of 8 different loss values—can be made from these sets. Which lotteries participants saw, and what task they performed when they saw different lotteries—a decision about whether to accept or pass on the gamble, or another task—varied by condition (see Figure 2):

*Figure 1. An example lottery used in Studies 1 and 2*
Figure 2. Summary of conditions used in Studies 1-2 to determine what places values in the decision sample. Each row reflects a condition that builds on the row below it. For example, those in the narrow + gain evaluation condition did not merely have to respond
Wide. These participants saw and made decisions about all 112 lotteries, presented in a random order. They indicated whether they would accept or reject each lottery.

Narrow + Evaluation. Whereas participants in this condition also saw all 112 lotteries, they only made a decision about whether to accept or reject lotteries over a narrow gain range: those that offered the chance to win between $6 and $20. For lotteries with gain values between $22 and $32, participants made a judgment that would require them merely to subjectively evaluate this value without making a decision about the gamble. More specifically, they rated the attractiveness of the gamble on a non-numeric slider scale anchored at not at all attractive and extremely attractive. The slider defaulted at the midpoint, which was labeled somewhat attractive.

Narrow + Response. Participants made decisions about the same 64 lotteries that those in the narrow + evaluation condition did. But for the other lotteries—those with gain values between $22 and $32—participants responded in light of the values, but not in a way that required them to form a subjective evaluation. Instead, participants merely typed the gain value.

Narrow + Exposure. Like in the other two narrow conditions, narrow + exposure participants made decisions about all lotteries with gains between $6 and $20. But in this condition, participants were only exposed to the gain values from $22 to $32 on the number line itself. That is, participants did not respond in light of the values. So that participants in this condition would confront roughly the same number of trials as participants in the other two conditions, participants saw the set of 64 narrow-gain-range lotteries twice. Like in all conditions, the order of the lotteries was randomized. But for this condition, such randomization was subject to one constraint: Identical lotteries appeared only once in each half. Given Supplemental Study B found that exposure to values (i.e., $22 through $32 on the gain number
line) is insufficient to place them into the decision sample, this condition serves as a baseline against which to test whether responding to, evaluating, and/or deciding based on values places them in the decision sample.

**Results**

To determine which values entered the decision sample, we tested how our manipulations influenced participants’ degree of loss aversion. First, we had to calculate participants’ loss aversion coefficients. For each participant, we performed a logistic regression in which we predicted a particular participant’s decision to accept (+1) or reject (-1) a lottery as a function of its gain value and its loss value. From this regression, we took the beta for the loss value and divided it by the beta for the gain value. This quotient, once multiplied by negative one, reflects participants’ relative sensitivity to losses vs. gains when considering risky decisions. Next, we followed Walasek and Stewart’s (2015) three-step exclusion criteria by omitting those: with missing responses, whose regression fit deviance scores were among the remaining highest 5%, and whose loss aversion coefficients were negative (thereby displaying an inexplicable preference for lower-expected-value lotteries.) After exclusions, 993 participants remained. The interested reader can find the results without exclusions reported in the Supplemental Materials.

We began by comparing our two most-different conditions. As DbS anticipates, those who made decisions over the wide range of gains ($2 to $32) displayed more loss aversion ($Median coefficient = 1.31; 95% CI = [1.20, 1.39]) than narrow + exposure participants who only made decisions over the narrow ($2 to $20) range of gains ($Median coefficient = 1.02; 95% CI = [1.00, 1.06]), Z = 4.77, p < .001. Wide gain range participants had to make decisions over, subjectively evaluate, and respond in light of additional values. Which process (or processes) pushed those additional values into the decision sample?
To probe this question, we next examined the loss aversion coefficient in the narrow + response condition. If responding in light of values (instead of merely being exposed to them) pushes values into the decision sample, narrow + response participants should show more loss aversion than narrow + exposure conditions. They did not. Instead, narrow + response participants actually showed a marginal decline in loss aversion (Median coefficient = 1.00; 95% CI = [1.00, 1.05]), Z = -1.89, p = .058.

Next, we compared the narrow + evaluation condition to both of the just-reviewed conditions. If values that feed into subjective valuations are placed into the decision sample, then we would expect a greater loss aversion coefficient in this condition. That is what we observed. Evaluation produced a greater loss aversion coefficient (Median coefficient = 1.10; 95% CI = [1.07, 1.21]) than both the narrow + response condition, Z = -3.74, p < .001, as well as the narrow + exposure condition, Z = -2.23, p = .026. This supports that subjective evaluation is sufficient to place values into the decision sample (see Figure 3).

Does actually making decisions in light of a value—an act that also requires exposure, response, and evaluation—further increase the value’s tendency to enter the decision sample? At first glance, the answer seemed to be yes: We found that the wide condition produced an even higher loss aversion coefficient than the narrow + evaluation condition, Z = 2.43, p = .015. But note that this comparison does not provide clear support for this decision account: Instead, it may be that the loss aversion coefficient is greater when calculated over a different set of lotteries—a set that also includes those with higher gains (those from $22 to $32) and thus higher expected values. To make the comparison more comparable, we recomputed a loss aversion coefficient for those in the wide condition using only those lotteries with gains ranging from $6 to $20. In this case, the wide condition produced no greater loss aversion (Median coefficient = 1.19; 95% CI =
**Figure 3.** The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition (Study 1).
EVALUATIONS ARE COMPARATIVE, BUT TO WHAT?  13

[1.10, 1.33]) than the narrow + evaluation condition, $Z = 1.33, p = .185$. This suggests that actually make decisions did not contribute to placing values in the decision sample; only evaluation had this effect.

**Study 2**

Although Study 1 identified subjective evaluation as the crucial process that places attribute values into a decision sample, Study 2 answers the question of precisely what must be evaluated. In Study 1’s evaluation condition, participants were asked to subjectively evaluate the lottery, which presumably required them to evaluate that lottery’s components (e.g., the gain and loss values). Study 2 tested whether it would be sufficient for participants to evaluate merely the gain value itself in order for it to enter the decision sample and guide decision making.

**Method**

**Participants and design.** Participants were 1,890 Americans recruited via AMT. Participants were randomly assigned to one of four gain range conditions: narrow + exposure, narrow + gain evaluation, narrow + lottery evaluation, or wide.

**Procedure.** The procedure was similar to the one used in Study 1. The narrow + exposure, narrow + lottery evaluation, and wide conditions were all repeated from the previous study. The difference between the new narrow + gain evaluation and the previously used narrow + lottery evaluation condition is what participants subjectively evaluated. In both conditions, participants offered evaluations related to lotteries whose gains ranged from $22 to $32. Whereas those in the narrow + lottery evaluation condition evaluated the lotteries, those in the new narrow + gain evaluation condition rated the attractiveness of the possible gain. These ratings were collected on the same unnumbered slider scale from Study 1, anchored at *not at all* and *extremely*
attractive. The midpoint—where by default the slider began—was again labeled somewhat attractive.

Results

We calculated loss aversion coefficients and trimmed the data using the same procedures described previously. One hundred fifteen participants were excluded based on Walasek and Stewart’s (2015) exclusion criteria. An additional 496 participants failed to answer an attention check question accurately. This left 1,279 participants for the final analysis. See the Supplemental Materials for the attention check question and the analyses without exclusions.

As shown in Figure 4, we again found that participants displayed greater loss aversion when they made decisions over a wide range of gains (Median coefficient = 1.47, 95% CI = [1.30, 1.59]) compared to a narrow range of gains (Median coefficient = 1.07; 95% CI = [1.03, 1.10]), Z = 6.88, p < .001. Furthermore, we replicated the findings from Study 1 that evaluating the attractiveness of the wider range of lotteries (narrow + lottery evaluation) also elevated loss aversion (Median coefficient = 1.26; 95% CI = [1.17, 1.33]) compared to the narrow gain range condition, Z = 4.21, p < .001.

Participants in the new narrow + gain evaluation condition showed elevated loss aversion (Median coefficient = 1.29; 95% CI = [1.16, 1.41]) compared to the narrow + exposure condition, Z = 4.56, p < .001. Furthermore, the two evaluation conditions were not statistically distinguishable, Z = .40, p = .69. In other words, evaluation—whether of the lottery (meaning the gain and loss together) or the gain value directly—was sufficient to place gain values in the decision sample.

Finally, we considered the role of actually making decisions based on the values (as opposed to merely evaluating them) by comparing the two evaluation conditions to the wide gain
**Figure 4.** The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition (Study 2).
range condition. Those in the wide condition showed marginally greater loss aversion than those in the narrow + gain evaluation condition, $Z = 1.68, p = .094$, and significantly more than the narrow + lottery evaluation, $Z = 2.23, p = .025$. But recall a concern raised in Study 1: Did this elevated loss aversion reflect the influence of making a decision on placing values in the decision sample, or did it reflect that wide condition participants’ loss aversion coefficient was calculated over a different sample of lotteries? To disentangle these possibilities, we recalculated the loss aversion coefficients using only those lotteries that all participants accepted or rejected—i.e., those with gain values ranging from $6$ to $20$. This reduced the loss aversion observed in the wide condition ($\text{Median coefficient} = 1.24; 95\%\ CI = [1.16, 1.44]$) so that it was no longer greater than those in the narrow + lottery evaluation, $Z = -1.05, p = .292$, or the narrow + gain evaluation conditions, $Z = -1.54, p = .123$. In summary, Study 2 provides firmer support that it is the evaluation of values—not actually making a decision—that is responsible for placing those values in the decision sample.

**Study 3**

Our aim is to understand when values enter a decision sample—not merely to influence loss aversion, but judgment and decision making more generally. Consistent with this goal, Study 3 extended our investigation to a new domain: patience. Participants considered tradeoffs between smaller-sooner and larger-later monetary payouts. We varied the set of temporal (instead of monetary) values that might enter participants’ decision samples and affect their willingness to delay payoffs. People’s willingness to accept delays should depend on how subjectively short they seem. Although all participants made choices over the same set of tradeoffs, they either subjectively evaluated or merely responded to (i.e., retyped) other temporal values that came from a uniform or skewed (with many near times and few distal times)
distribution. If evaluation introduces these values into the decision sample, then the nature of the distribution should have a stronger influence on participants’ display of patience in a way that DbS would expect.

**Method**

**Participants and design.** We recruited 1,209 Americans from AMT. Participants were randomly assigned to one of four conditions in a 2(time distribution: uniform or skewed) X 2(task: response or evaluation) full-factorial design. We excluded 229 participants who were unable to answer a simple attention check correctly (see the Supplemental Materials). All analyses were conducted on the remaining 980 participants.

**Procedure.** Participants learned they would see 60 pairs of payoffs. Each pair would feature a certain amount of money that could be received immediately or a larger amount of money that would be received after a certain time delay. The time distribution manipulation determined whether those time delays came from a highly skewed distribution (1 day, 1 week, 1 month, 2 months, 6 months, 12 months) or a relatively uniform distribution (2 months, 4 months, 6 months, 9 months, 12 months, 15 months.) Note that three of the time delays are common to the conditions: 2 months, 6 months, and 12 months. Crucially, the ranks of these values in each condition were either highly discrepant (2 months: 3rd vs. 6th), somewhat discrepant (6 months: 2nd vs. 4th), or barely discrepant (12 months: 1st vs. 2nd). Or considered differently, the three delays are more similar in rank in the skewed distribution (occupying the adjacent 1st, 2nd, and 3rd positions), whereas these same delays have more differentiated ranks in the relatively uniform distribution (occupying the non-adjacent 2nd, 4th, and 6th positions).

The immediate payoffs were $100, $200, $300, or $400. The delayed payoffs were $200, $300, $400, or $500. We defined tradeoffs using only those 10 combinations in which the
delayed payoff would be greater than the immediate payoff. And because in each condition the time delay for the larger payoff could take one of 6 distinct forms, this means each participant saw 60 tradeoffs. Thirty of the pairs were common to both conditions. (These were the ones that involved delays of 2, 6, or 12 months.) On these choice trials, participants indicated whether they would prefer the smaller amount today or the larger amount after the specified delay.

But for the other thirty pairs—those that involved delays that were unique to that particular distribution condition—participants did not indicate their preference. Instead, what they did depended on their task condition. In the response task condition, participants were asked to “type the amount of time you will have to wait to receive the larger payout.” In the evaluation task condition, participants were asked to indicate “how unappealing it would be to have to wait [time delay] for the payoff?” Participants responded on a slider scale that ranged from “not at all unappealing” to “extremely unappealing.” The middle was labeled “somewhat unappealing.”

To make sure that responding to (retyping) or evaluating values had a chance to modify participants’ decision samples before their very first choice trials, we had all participants consider these three time delays—i.e., those specific to each condition and that would be used on the subsequent response or evaluation trials—before beginning the main task. Response participants had to retype the three time delays. Evaluation participants had to indicate how unappealing it would be to wait those amounts of time for a reward of $200 to $500. At that point, the 60 tradeoffs appeared in random order.

Results

We began by testing whether those exposed to the uniform distribution of time delays displayed more patience than those who saw the skewed distribution. After all, the required delays for the larger reward should have seemed subjectively shorter when considered in the
context of the uniform than the skewed distribution. And indeed, that was the case. In this and all
subsequent models, we used a mixed model with fixed effects predictors and a random effect of
participant (to account for the non-independence of each participant’s 30 choices). When
considering the same tradeoffs, uniform participants indicated a willingness to wait for the larger
reward on more trials (55.35%) than did skewed participants (50.50%), \( t(976.98) = 2.41, p
= .016. \)

Did such patience take the form that DbS would anticipate? More specifically, we
expected that uniform participants’ greater patience would emerge most clearly at 2 months
(when the rankings were highly discrepant between conditions) and least strongly at 12 months
(when the rankings were barely discrepant between conditions). We coded time delay as -1 (2
months), 0 (6 months), and +1 (12 months). As expected, we observed a Time Distribution X
Time Delay interaction, \( t(28419) = 8.12, p < .001 \). This reflected the expected pattern. Uniform
participants were much more patient at 2 months, \( t = 4.34, p < .001 \); less so at 6 months, \( t = 2.41,
\( p = .016; \) and did not differ significantly from skewed participants at 12 months, \( t < 1. \)

To probe our central question—whether evaluation places values in the decision
sample—we tested whether this finding was driven by participants in the evaluation (compared
to the response) condition. And indeed, the Time Distribution X Time Delay X Task interaction
was significant, \( t(28414.99) = 5.03, p < .001 \). When participants subjectively evaluated the
additional values, a significant Time Distribution X Time Delay interaction suggested that such
values populated the decision sample and influenced patience as decision by sampling would
predict, \( t(28414.99) = 10.01, p < .001 \). In contrast, when participants merely retyped the
additional values, this Time Distribution X Time Delay interaction was much weaker,
\( t(28414.99) = 2.51, p = .012 \) (see Figure 5). Summarized differently, although there was some
evidence that merely retyping values placed values in the decision sample, subjectively evaluating those values produced evidence of this that was almost four times as strong.

**Figure 5.** Proportion of participant choices reflecting patience (choice of the larger-later option) as a function of task condition and delay. Error bars reflect ±1 standard error from the mean. (Study 3).

**General Discussion**

Behavioral scientists have struggled for decades to understand how people characterize the magnitude of quantitative attributes. Especially considering psychologists’ longstanding interest in this topic, DbS offers a relatively new answer. DbS moves beyond a mere description of the relationship between objective quantities and subjective valuations to posit that such valuations are arrived at through a comparative process. The theory posits that people subjectively assess attribute values by comparing them to a sample of numbers drawn from memory.

“We assume that the decision sample, to which a target…is compared, is a small, random sample…from memory” (Stewart et al., 2006, p. 4). The authors go on to say that “of course this random sampling assumption is likely to be incorrect” (p. 4). Subsequent research examined
several properties of attribute values that make them more or less likely to enter decision samples. The present paper instead examined whether how people *process* values affects whether they enter a decision sample. By answering this complementary question, the ambitious goals and full potential of DbS be more fully realized.

We found that neither being exposed to nor having to respond to values is sufficient to place them in the decision sample. In other words, manipulations that merely make values accessible—through mere exposure or a more involved, response-oriented consideration—do not consistently push them to be comparison standards that guide subjective valuation. Instead, subjectively *evaluating* values led them to enter that pool used to subjectively evaluate additional ones.

The present paper provides a preliminary, but certainly not a complete, understanding of the decision sample. Although evaluation inserts values into that sample, it remains unclear just how long they remain. After all, with time, a person will be exposed to and evaluate new values. To what extent do such new values merely complement or instead displace values already residing in the sample? Instead of asking what leads numbers to enter into a decision sample, future research could ask what leads certain numbers to depart the decision sample. One possibility is that these values’ membership in the sample simply fades with time. Another possibility is that the depth with which these values were evaluated—whether the values were subject to a cursory assessment or a more thorough analysis—may determine their longevity.

To apply the present findings to new real-world judgment and decision making contexts, more research must be done to understand which attribute values are *spontaneously* evaluated. That is, in naturalistic contexts, people are not confronted with experimental manipulations that ask them to subjectively characterize values. Consider the finding that cultures differ in their
reactions to death depending on the distribution of death tolls to which they are exposed by the media (Olivola & Sagara, 2009). It certainly does seem intuitive—and if the present paper is correct, it should be the case—that quantifiable tragedies, perhaps in part out of empathy, are events whose scope people spontaneously evaluate. Characterizations of mass casualties as “unprecedented in number” or more limited tragedies as those that “certainly could have been worse” reflect the sort of subjective assessments that should place those values into decision samples. To continue with this example, when might spontaneous evaluation not occur? Although local media are disproportionately likely to consider local events, people also learn of global tragedies. One possibility is that local tragedies—given they may prompt more interest and concern—are more likely to be subjectively evaluated than global ones.

In conclusion, psychologists have long appreciated that people make sense of the world based on their context. Decision by sampling formalizes how people both make sense of and rely on attribute values to guide judgment and decision making. Understanding which context cues guide this process requires moving beyond determining to which values people are merely exposed to instead learn which values people spontaneously evaluate.
References


EVALUATIONS ARE COMPARATIVE, BUT TO WHAT?  24


SUPPLEMENTAL MATERIALS

Study A

In order to test what is responsible for placing values in a decision sample, we must be confident that we are leaning on a paradigm that does indeed test DbS’s influence. That is, we wanted to be confident that the loss aversion coefficient does indeed show variation in a way that DbS would anticipate. Walaesk and Stewart (2015) were able to increase or decrease loss aversion by changing the distribution of gains and losses that defined the set of lotteries to which participants were exposed. Gains and losses were distributed over a coarsely differentiated wide range ($12, $16, $20, …, $36, $40) or a finely differentiated narrow range ($6, $8, $10, …, $18, $20). When values varied over the finely differentiated narrow range, participants became more sensitive to variation on that dimension. In other words, loss aversion was strongest when losses varied from -$20 to -$6 in increments of $2 and gains varied from $12 to $40 in increments of $4.

Let us consider more deeply why DbS anticipates these results. DbS essentially suggests that values are rescaled or normalized according to their ranking. That is, in moving from values at the lower end of a range (e.g., -$20 or $12) to the upper end (-$6 or $40), one should feel that the same subjective distance has been traversed. As long as the values in between are evenly spaced out (in $2 or $4 increments), then the wider the range of the decision sample, the less sensitive one should become to the same objective change on that dimension.

But is it important that the manipulation of range is confounded with granularity—i.e., whether the values differ by $2 or $4? According to DbS, at least for this particular loss aversion paradigm, no. Whether the wide range of values ($12 to $40) is accompanied by more middling numbers that are $2 or $4 apart (i.e., $12, $14, $16, …, $40 vs. $12, $16, $20, … $40), it is still
the case that moving from the lower end to the midpoint of the decision sample involves the same change in value (i.e., $14) and the same change in normalized rank (from 0 to .5) that DbS identifies as crucial. Considered differently, each $4 shift may produce a change in two ranks instead of one, but with twice as many distinct values in the finely differentiated distribution, there is no net effect in what DbS predicts such a shift would feel like subjectively. In other words, if this paradigm does probe the operation of the decision sample, then the loss aversion coefficient should be sensitive to the values’ range, not how finely differentiated the values are along that range.

An alternative possibility is DbS does not offer the best explanation for Walasek and Stewart’s (2015) findings. That is, perhaps the confounding feature—interval value—explains the results. When the values differed in a more fine-grained manner, decision makers may have learned to make more subtle differentiations between them. If this, not DbS, explains Walasek and Stewart’s (2015) findings, then it suggests that exposing people to a wide range of finely differentiated values should also sensitize people to changes on that dimension. Resolving such ambiguity is critical if we wish to lean on this paradigm to address bigger questions about decision by sampling.

Study A distinguished between these two possibilities by unconfounding the range of values from the granularity of the interval value intervals. We randomly assigned participants to be exposed to a wide range of gains ($12 to $40) or a wide range of losses (-$40 to -$12). The other dimension was always narrowly defined ($6 to $20, or -$20 to -$6). For some participants, we replicated Walasek and Stewart’s paradigm by keeping the interval value intervals asymmetric—$4 for the wide range and $2 for the narrow range. But for other participants, we allowed the interval to be symmetric—$2 for both ranges. If DbS explains fluctuations in loss aversion, we
should see only a main effect of range: Loss aversion should be higher when considering a wide range of gains instead of a wide range of losses. But if sensitivity to the granularity with which values vary explains the degree of loss aversion, then eliminating the range-granularity confound should disrupt this effect.

**Method**

**Participants and design.** We recruited 1,206 Americans from Amazon Mechanical Turk (AMT). Given the study included a partial replication of Walasek and Stewart (2015), we predetermined this sample size by following recommendations of Simonsohn (2015). We took Walasek and Stewart’s largest sample size (Study 1b, N = 423, also four conditions) and multiplied that value by 2.5. Participants were randomly assigned to one of four conditions in a 2(wide range: gain or loss) X 2(interval: asymmetric or symmetric) full-factorial design.

**Procedure.** We began by explaining to participants that they would be exposed to a series of lotteries. To reinforce that the chance of winning or losing each lottery was equivalent, we explained that each lottery’s outcome would be determined by the flip of a coin. Heads would produce a win; tails, a loss. Participants saw one example lottery (Figure S1), which illustrated the format that would be used on all of the main trials.

*Would you accept or reject the following 50/50 gamble?*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>-$12</strong></td>
<td><strong>+$6</strong></td>
</tr>
</tbody>
</table>

**Accept**

**Reject**
Participants were exposed to every lottery that could be defined by the specific range and interval conditions to which they were assigned (see Figure S2 for a summary). This meant that those in the symmetric interval conditions actually responded to a larger number of lotteries (120) than did those in the asymmetric interval conditions (64). But to roughly equate the number of trials participants completed, those in the asymmetric interval condition saw each distinct lottery twice. Subject to the constraint that repeat lotteries had to appear in different halves of the sequence of lotteries, the lotteries appeared in a randomly determined order.

Results and Discussion

To determine whether our manipulations influenced participants’ degree of loss aversion, we first had to calculate participants’ loss aversion coefficients. For each participant, we conducted a logistic regression in which we predicted a particular participant’s decision to accept (+1) or reject (-1) a lottery as a function of the gain value and the loss value of the lottery. From this regression, we took the beta for the loss value and divided it by the beta for the gain value. This value, once multiplied by negative one, reflects participants’ relative sensitivity to losses vs. gains when considering risky decisions (see Walasek & Stewart, 2015).

In this and every study we precisely followed Walasek and Stewart’s (2015) exclusion criteria, which we describe next. First, we excluded participants with incomplete responses. Second, we excluded those participants whose regression fit deviance scores were among the remaining highest 5%. Third, we excluded those who displayed a negative loss aversion coefficient. Such participants indicated greater interest in lotteries with lower expected values, indicating a failure to understand the procedure or take it seriously. After exclusions, we had 868 participants remaining.
Given the loss aversion coefficients were not distributed normally, we tested our
Table 2. The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition (Study A).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Range of Gains for Lottery Decisions</th>
<th>Range of Losses for Lottery Decisions</th>
<th>Interval Between Gain Values</th>
<th>Interval Between Loss Values</th>
<th>Median Loss Aversion Coefficient with Bootstrapped 95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow - Asymmetric Interval</td>
<td>[6, 20]</td>
<td>[-40, -12]</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Wide – Asymmetric Interval</td>
<td>[12, 40]</td>
<td>[-20, -6]</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Narrow-Symmetric Interval</td>
<td>[6, 20]</td>
<td>[-40, -12]</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Wide – Symmetric Interval</td>
<td>[12, 40]</td>
<td>[-20, -6]</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Figure S2: The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition (Study A).
hypotheses using 4 pairs of nonparametric tests. The asymmetric interval conditions replicated Walasek and Stewart (2015). That is, the loss aversion coefficients were higher when gains were distributed over a wide range (Median coefficient = 1.44) compared to when losses were (Median coefficient = 0.83), Z = 12.81, p < .001. As DbS anticipates, we found that even when the intervals were made symmetric, loss aversion continued to be higher when gains had a wide range (Median coefficient = 1.44) than when losses did (Median coefficient = 0.89), Z = 12.45, p < .001.

Did we find any evidence that making the interval symmetric for gains and losses did anything to reduce the basic effect? In short, no. When participants were exposed to a wide range of gains, the loss aversion coefficient remained just as high when the interval was reduced from $4 to $2, Z = 0.52, p = .607. And when participants were exposed to a wide range of losses, the loss aversion coefficient remained just as low when the interval was reduced from $4 to $2, Z = 0.44, p = .659.

**Study B**

Study B tested whether exposure was sufficient to place values into the decision sample. All participants saw lotteries defined over the same narrow range of losses (-$20 to -$6). What we varied was the range of gains to which participants were exposed as well as whether participants actually made decisions over the same range of gain values. Those in the **wide** condition made decisions about lotteries whose gains ranged from $6 to $32. Those in the **narrow** condition made decisions about lotteries whose gains ranged from $6 to $20.

DbS is clear that the loss aversion coefficient should be higher in the wide condition compared to the narrow condition. But is this merely (or partly) because those in the wide condition were *exposed* to a wider range of gain values? A third condition was instrumental to
answering that question. In a narrow + exposure condition, participants made decisions about lotteries over the narrow gain range ($6 to $20) but were exposed to the full range of gain values ($6 to $32). If exposure is sufficient to place values into the decision sample, then the narrow + exposure condition should prompt a higher loss aversion coefficient than the narrow condition. And if exposure is the sole determinant of what enters the decision sample, then this elevated loss aversion should be statistically indistinguishable from the elevated loss aversion coefficient of the wide condition. If instead exposure is not sufficient to insert values into the decision sample, the loss aversion coefficient for those in the narrow + exposure condition should be similar to those in narrow condition (and thus smaller than those in the wide condition).

Method

Participants and design. Participants were 1,070 Americans recruited through AMT. They were randomly assigned to one of three gain range conditions: wide, narrow, or narrow + exposure. Although this design included one fewer condition than Study A, we were testing whether the basic effect documented in that study may be decomposable into component parts. Also, our wide range was slightly narrower than that used in Study A. At the same time, the strength of our results in Study A suggested it was well powered. Combining these observations led us to aim for a roughly equivalent sample size as in Study A.

Procedure. The procedure was similar to that used in Study A except for the following changes. First, all participants saw lotteries defined by the same narrow range of losses (-$20 to -$6, in $2 increments). Second, we modified the way that lotteries were presented in order to decouple the range of values used to define lotteries from the range of values to which participants were exposed. On every trial, participants were exposed to the same two number
lines—one for gains, one for losses. The gain and loss values for a particular lottery were identified on their respective number line (see Figure S3).

We varied the range of values to which participants were exposed by varying the width of the number line. The gain number line spanned from $6 to $20 in the narrow condition (Figure S3A), but from $6 to $32 in the wide condition (Figure S3B). Although the gain values of the lotteries in the narrow + exposure condition ranged only from $6 to $20 (as in the narrow condition), the number line ranged from $6 to $32 (as in the wide condition). All participants saw a loss number line that ranged from -$20 to -$6.

Those in the wide condition indicated whether they would accept or pass on the 112 unique lotteries that could be defined by every combination of the 14 gain and 8 loss values. In contrast, those in the narrow and narrow + exposure conditions saw the 64 unique lotteries that could be created by every combination of the 8 gain and 8 loss values. In order to roughly equate across the number of lotteries that participants saw, these participants responded to these 64 lotteries twice (though always in different halves of the trials). Subject to this one constraint, the order of the lotteries was randomized.

Results and Discussion

We computed a loss aversion coefficient for each participant in an equivalent way as was done in Study A. At that point, we followed the same steps outlined in Study 1 to identify and exclude outlier and non-compliant participants. This left 1,002 participants in all analyses reported below. Conceptually replicating Study A, we found that participants exposed to a wide range of gains showed greater loss aversion (Median coefficient = 1.13) compared to those who saw a narrow range of gains (Median coefficient = 1.03), Z = 3.35, p < .001.

Was exposure to a wide range of gains sufficient to place those values in the decision
sample, thereby inflating the loss aversion coefficient? In a word, no. Those in the narrow +
exposure condition showed a relatively low loss aversion coefficient (Median = 1.02), roughly
comparable in size to that of the narrow condition, $Z = 0.95, p = .343$. This loss aversion
Figure S3. An example lottery as seen by those in the the narrow condition (A) and wide and narrow + exposure conditions (A) in Study B.
coefficient was significantly smaller than shown by those who were not merely exposed to but actually made lottery decisions over the wide range of gains, $Z = 4.16, p < .001$ (see Figure S4).

**Additional Analyses**

In this section, we report analysis including all participants in Studies A-B, 1-3. Where appropriate, we also report analyses including decisions only on the $6 - $20 lottery gain range to make a cross-condition comparison even more parallel (Studies 1 and 2). These analyses are more conservative because they compute the loss aversion coefficient over the same set of lotteries for all participants, thereby allowing us to focus on how our manipulations—not differences in the lotteries used to compute the loss aversion coefficients—change participant’s loss aversion.

**Study A**

**Analysis including all participants.** When the intervalue intervals were asymmetric, we found that loss aversion was higher when gains were distributed over a wide range ($Median \ coefficient = 1.39$) compared to when losses were ($Median \ coefficient = 0.77$), $Z = 12.62, p < .001$. But even when the intervals for gains and losses were made symmetric at $S2$, the loss aversion coefficient was higher when gains had a wide range ($Median \ coefficient = 1.39$) than when losses did ($Median \ coefficient = 0.80$), $Z = 12.61, p < .001$. The loss aversion coefficients between the symmetric and asymmetric conditions did not differ for participants exposed to the narrow range of gains, $Z = .07, p = .944$, or the wide range of gains, $Z = .49, p = .626$ (See Figure S4).

**Study B**
**Analysis including all participants.** Participants exposed to a wide range of gains showed greater loss aversion (*Median coefficient* = 1.12) compared to those who saw a narrow range of gains (*Median coefficient* = 1.01), $Z = 3.60, p < .001$.

Those in the narrow + exposure condition showed a relatively low loss aversion
Figure S4. The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition including all participants (Study A).
coefficient (Median coefficient = 1.01), roughly comparable in size to that of the narrow condition, \( Z = .691, p = .489 \). This loss aversion coefficient was clearly smaller than the coefficient of those in the wide condition—participants who actually made decisions about lotteries defined over the wider range of gains, \( Z = 4.24, p < .001 \) (See Figure S5).

**Study 1**

**Analysis including all participants.** Participants in the wide condition showed more loss aversion (Median coefficient = 1.26) than those in the narrow + exposure condition (Median coefficient = 1.01), \( Z = 4.69, p < .001 \).

Participants in the narrow + response condition showed a similar level of loss aversion (Median coefficient = 1.00) as those in the narrow + exposure condition (Median coefficient = 1.10), \( Z = 1.04, p = .299 \), and significantly less than those in the wide condition, \( Z = 5.32, p < .001 \).

Participants in the narrow + evaluation condition had a higher loss aversion coefficient (Median coefficient = 1.10) than those in the narrow + response condition, \( Z = 3.26, p = .001 \), or those in the narrow + exposure condition, \( Z = 2.51, p = .012 \), albeit still less loss averse than those in the wide condition (Median coefficient = 1.31), \( Z = 2.16, p = .030 \) (see Figure S6).

**Analysis including decisions only on the S$6 - S$20 lottery gain range.** We conducted the same analysis with participants’ lottery decisions on the same limited range but without excluding any participants. Participants in the wide condition showed more loss aversion (Median coefficient = 1.16) than those in the narrow + exposure condition (Median coefficient = 1.01), \( Z = 3.83, p < .001 \).
Participants in the narrow + response condition displayed similar loss aversion (*Median coefficient* = 1.00) to those in the narrow + exposure condition, *Z* = 1.04, *p* = .299, and significantly less loss aversion than those in the wide condition, *Z* = 4.56, *p* < .001.
Figure S5. The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition including all participants (Study B).
Figure S6. The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition including all participants (Study 1).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Range of Gains for Lottery Decisions</th>
<th>Range of Gains for All Lotteries</th>
<th>Range of Losses</th>
<th>Interval Between Values (Losses, Gains)</th>
<th>Non-Decision Task</th>
<th>Median Loss Aversion Coefficient with Bootstrapped 95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wide</td>
<td>[$6, $32]</td>
<td>[$6, $32]</td>
<td>[-$20, -$6]</td>
<td>S2</td>
<td>N/A</td>
<td><img src="graph.png" alt="Graph" /></td>
</tr>
<tr>
<td>Narrow + Lottery Evaluation</td>
<td>[$6, $20]</td>
<td>[$6, $32]</td>
<td>[-$20, -$6]</td>
<td>S2</td>
<td>evaluate attractiveness of lottery</td>
<td><img src="graph.png" alt="Graph" /></td>
</tr>
<tr>
<td>Narrow + Response</td>
<td>[$6, $20]</td>
<td>[$6, $32]</td>
<td>[-$20, -$6]</td>
<td>S2</td>
<td>respond to gain</td>
<td><img src="graph.png" alt="Graph" /></td>
</tr>
<tr>
<td>Narrow + Exposure</td>
<td>[$6, $20]</td>
<td>[$6, $32]</td>
<td>[-$20, -$6]</td>
<td>S2</td>
<td>N/A</td>
<td><img src="graph.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
Participants in the narrow + evaluation condition had a higher loss aversion coefficient (Median coefficient = 1.10) than those in the narrow + response condition, $Z = 3.26, p = .001$, or those in the narrow + exposure condition, $Z = 2.51, p = .012$, and similar loss aversion to those in the wide condition (Median coefficient = 1.16), $Z = 1.54, p = .123$.

**Study 2**

**Attention check question.** The attention check question was, “Different participants are asked to do different things in this study. Which most accurately describes what you were asked to do?” Participants selected one from the following four options. The accurate answer depended on the condition participants were randomly assigned to:

- For all lotteries, you indicated whether you would accept or reject them.
- For some lotteries (but not all), you rated the attractiveness of the outcome if the lottery came up heads (instead of the lottery as a whole)
- For some lotteries (but not all), you rated the attractiveness of the lottery as a whole.
- For some lotteries (but not all), you rated the attractiveness of the outcome if the lottery came up tails (instead of the lottery as a whole).

**Analysis including all participants.** Participants were more loss averse in the wide condition (Median coefficient = 1.44) than in the narrow + exposure condition (Median coefficient = 1.05), $Z = 7.63, p < 0.001$. Furthermore, participants who subjectively evaluated lotteries (i.e., the narrow + lottery evaluation condition) showed more loss aversion (Median coefficient n = 1.20) than those in the narrow + exposure condition, $Z = 4.52, p < 0.001$, albeit less than in the wide condition, $Z = 3.20, p = .001$.

Participants in the narrow + gain evaluation condition were similarly loss averse to those in the wide condition (Median coefficients: 1.28 vs. 1.44), $Z = 1.74, p = .082$, and those in the
narrow + lottery evaluation condition, $Z = 1.39, p = .164$, but substantially more loss averse than those in the narrow + exposure condition, $Z = 5.90, p < .001$ (See Figure S7)
**Figure S7.** The characteristics of the lotteries and the median loss aversion coefficients, complete with 95% bootstrapped confidence intervals, by condition including all participants (Study 2).
Analysis including decisions only in the $6 - $20 gain range and including those who failed the attention check. We again conducted analyses on participants’ lottery decisions on the same limited range but without excluding any participants. Participants were more loss-averse in the wide condition (Median coefficient = 1.20) than in the narrow + exposure condition (Median coefficient = 1.05), $Z = 3.01, p = 0.003$. Furthermore, participants asked to subjectively evaluate lotteries (i.e., the narrow + lottery evaluation condition) showed more loss aversion than those in the narrow + exposure condition (Median coefficient = 1.20), $Z = 4.63, p < 0.001$, and similarly loss averse as those in the wide condition, $Z = .99, p =.321$.

Participants in the narrow + gain evaluation condition were surprisingly more loss averse than those in the wide condition (Median coefficients: 1.29 vs. 1.20, $Z = 2.53, p =.012$), similarly loss averse than those in the narrow + lottery evaluation condition (Median coefficient = 1.19), $Z = 1.47, p =.142$, and substantially more loss averse than those in the narrow + exposure condition, $Z = 6.51, p <.001$. Perhaps directly prompting people to evaluate the gains is the most direct assurance that such evaluation happens, thereby most efficiently placing those values in the decision sample.

Study 3

Attention check question. In this study, you saw 60 pairs of payoff options. Different participants are asked to do a different task with these pairs of payoff options. Which most accurately describes what you were asked to do?

- For all pairs of payoff options, I made choices about which of one of the two options I prefer.
- For some pairs of payoff options, I made choices about which one of the two I prefer, but for the other pairs, I typed in the lengths of time I would have to wait to receive the larger payoff.
in a blank space

○ For some pairs of payoff options, I made choices about which one of the two I prefer, but for the other pairs, I indicated how unappealing it would be to wait for a certain amount of time to receive the larger payoff.

○ For some pairs of payoff options, I typed in the lengths of time I would have to wait to receive the larger reward in a blank space, but for the other pairs, I indicated how unappealing it would be to wait for a certain amount of time to receive the larger payoff.

**Analysis including all participants.** We began by testing whether those exposed to the uniform distribution of time delays displayed more patience than those who saw the skewed distribution. After all, the required delays for the larger reward should have seemed subjectively smaller when considered in the context of the uniform than the skewed distribution. And indeed, that was the case. In this and all subsequent model, we used a mixed model with fixed effects predictors and a random effect of participant (to account for the non-independence of each participant’s 30 choices). When considering the same tradeoffs, uniform participants indicated a willingness to wait for the larger reward on more trials (52.6%) than did skewed participants (48.4%), \( t(1205.96) = 2.28, p = .023 \).

Did such patience take the form that DbS would anticipate? More specifically, we expected that uniform participants’ greater patience would emerge most clearly at 2 months and least strongly at 12 months. We coded time delay as -1 (2 months), 0 (6 months), and +1 (12 months). As expected, we observed a Time Distribution X Time Delay interaction, \( t(35060) = 11.03, p < .001 \). This reflected the expected pattern. Uniform participants were much more patient at 2 months, \( t = 4.85, p < .001 \); less so at 6 months, \( t = 2.28, p = .023 \); and did not differ significantly from skewed participants at 12 months, \( t < 1 \).

To probe our central question—whether evaluation places values in the decision sample—we tested whether this finding was driven by participants in the evaluation (compared
to the response) condition. And indeed, the Time Distribution X Time Delay X Task interaction was significant, $t(35055.99) = 4.85, p < .001$. When participants subjectively evaluated the additional values, a significant Time Distribution X Time Delay interaction suggests that such values populated the decision sample and influenced patience as decision by sampling would predict, $t(35055.99) = 12.08, p < .001$. In contrast, when participants merely retyped the additional values, this Time Distribution X Time Delay interaction was much weaker, $t(35055.99) = 4.78, p < .001$. 