

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/141178>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

You Cannot Accurately Estimate An Individual's Loss Aversion Using An Accept-Reject  
Task

Lukasz Walasek

Neil Stewart

## Abstract

Loss aversion of the prospect theory is often measured in the accept-reject task, in which participants accept or reject the chance of playing a series of gambles. The gambles are two-branch 50/50 gambles with varying gain and loss amounts (e.g., 50% chance of winning \$20 and a 50% chance of losing \$10). Prospect theory quantifies loss aversion by scaling losses up by a parameter  $\lambda$ . Here we show that  $\lambda$  suffers from extremely poor parameter recoverability in the accept-reject task.  $\lambda$  cannot be reliably estimated even for a simple version of prospect theory with linear probability weighting and value functions.  $\lambda$  cannot be reliably estimated even in impractically large experiments with participants subject to thousands of choices. The poor recoverability is driven by a trade-off between  $\lambda$  and the other model parameters. However, a measure derived from these parameters is extremely well recovered—and corresponds to estimating the area of gain-loss space in which people accept gambles. This area is equivalent to the number of gambles accepted in a given choice set. That is, simply counting accept decisions is extremely reliably recovered—but using prospect theory to make further use of exactly which gambles were accepted and which were rejected does not work.

Would you play a gamble that offered the chance to win \$20 if a coin comes up heads but lose \$10 if the coin comes up tails? Suppose a person is indifferent—they do not mind whether they will or will not participate in this lottery. In such a case, we say that this person is exhibiting *loss aversion*. For this individual, the pleasure of receiving \$20 is equal, psychologically, to the pain of having to pay \$10. The idea that losses loom larger than gains is a cornerstone of behavioural economics. Loss aversion is often offered as the explanation for various phenomena in choice, like the endowment effect (Kahneman, Knetsch, & Thaler, 1991). Loss aversion is also a core component in prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), the leading descriptive model of choices where risks are involved. In this model, loss aversion is captured by the  $\lambda$  parameter.  $\lambda$  is the exchange rate between the psychological value of gains and losses. For example, with  $\lambda = 2$  (other things equal) one would be indifferent between accepting and rejecting the coin toss example—because  $20 - \lambda \times 10 = 0$ . Using people’s responses to a series of coin toss questions like the example here, where the gain and loss on offer in each toss are varied, is a popular method used to estimate a person’s  $\lambda$ . Here we show that most a commonly used modeling approach to estimate  $\lambda$  in this accept-reject task is inappropriate.

There is a growing literature questioning the theoretical and empirical aspects of loss aversion, in both risky and riskless context (Ert & Erev, 2013; Gal & Rucker, 2018; Yechiam, 2018). Rather than focusing on whether loss aversion exists, the present paper evaluates one of the commonly used methodologies of estimating at individual-level from risky choices. There are multiple approaches to estimating parameters of prospect theory, many of which have been designed with the main purpose of estimating individuals’ loss aversion (Abdellaoui, Bleichrodt, & L’Haridon, 2008; Rieskamp, Busemeyer, & Mellers, 2006; Stott, 2006; Pachur & Kellen, 2013). The *accept-reject* task—asking people to decide whether or not to play each in a series of coin-toss gain-loss questions like the one described above—is one popular approach. In a typical accept-reject experiment,

participants are presented with a series of mixed lotteries, each offering 50% chance of winning some amount, and a 50% chance of losing some amount of money. Individuals are simply required to indicate whether they would accept or reject each lottery. In the experiments that incentivize people's decisions, at least one of the lotteries is drawn at random at the end of the procedure to be played out for real outcomes.

From a practical and statistical point of view, the accept-reject task offers an elegant approach for determining how loss averse each participant is. First, the accept-reject task for estimating  $\lambda$  is very easy to administer. Since all lotteries offer two equally likely outcomes, the task can be explained with an analogy of a coin flip. Existing studies vary in the number of trials, but most range from 64 to 256. Thus, the accept-reject task requires relatively little cognitive effort and time from a participant. Indeed, this approach has been predominantly used in conjunction with more involved procedures, such as fMRI scanning. In one prolific study, the accept-reject task was used by Tom, Fox, Trepel, and Poldrack (2007) to show a positive correlation between activation in the dopaminergic regions of the brain and participants' behavioral loss aversion. Several authors relied on the accept-reject approach to further test the neural antecedents of loss averse behavior (Canessa et al., 2013; Chib, De Martino, Shimojo, & O'Doherty, 2012), or to show that individual-level variations in  $\lambda$  parameters correspond to differences in strategic thinking (Lorains et al., 2014), oxygen deprivation (Pighin, Bonini, Savadori, Hadjichristidis, & Schena, 2014), or the ability to accurately perceive internal physiological states (Sokol-Hessner et al., 2015). Table 1 below summarizes the key properties of 15 articles (from a meta-analysis of Walasek, Stewart, & Mullett, in prep) that used the accept-reject task to estimate loss aversion among their participants. The table shows that most studies find a considerable level of loss aversion, at least on the aggregate level.

The second benefit of using the accept-reject task is the simplicity with which participant's binary decisions can be converted into  $\lambda$  values. We illustrate the basic logic starting with the classic value function of prospect theory.

### Estimating Loss Aversion from Accept-Reject Choices

In cumulative prospect theory, the psychological value of a gain or loss  $x$  is given by

$$v(x) = x^\alpha \quad \text{if } x > 0 \quad (1)$$

$$v(x) = -\lambda|x|^\beta \quad \text{if } x \leq 0 \quad (2)$$

The curvature of the value function is determined by  $\alpha$  and  $\beta$ . Following Wakker (2010) and Stewart, Scheibehenne, and Pachur (2018) we set  $\alpha = \beta$ , because without this constraint  $\lambda$  is arbitrarily defined by the choice of unit for  $x$ , and does not measure the extent to which losses loom larger than gains across different values of  $|x|$ . Further, we set  $\alpha = \beta = 1$  to reduce the number of free parameters we must estimate from the accept-reject choice data. This is also useful because it is not possible to separately estimate loss aversion and risk aversion with a typical accept-reject task.

The prospect theory value of accepting a 50/50 gamble for +\$*gain* or -\$*loss* is given by

$$\begin{aligned} V(\text{accept}) &= w(1/2) v(\text{gain}) + w(1/2) v(-\text{loss}) \\ &= w(1/2) \text{gain}^\alpha - w(1/2) \lambda \text{loss}^\beta \end{aligned} \quad (3)$$

where  $w(1/2)$  is the decision weight attached to the 50/50 chance. Thus the first term is the subjective value of the 50% chance of winning *gain* and the second term is the subjective value of the 50% chance of losing *loss*. While cumulative prospect theory allows for the separate weighting of probabilities for gains and losses, we assume that  $w(1/2)$  is the same for the gain and the loss and, for the accept-reject task, this assumption is without loss of generality. This means that  $\lambda$  captures all loss aversion (see Pachur & Kellen, 2013).

To turn the subjective value of accepting the gamble into the probability of accepting the gamble we, just like other authors, use the exponentiated form of Luce's choice rule

(e.g., Glöckner & Pachur, 2012; Rieskamp, 2008; Scheibehenne & Pachur, 2015; Stewart et al., 2018; Stott, 2006). The probability of accepting the gamble is given by

$$p(\textit{accept}) = \frac{e^{bias} e^{\theta V(\textit{accept})}}{e^{bias} e^{\theta V(\textit{accept})} + e^{\theta V(\textit{reject})}} \quad (4)$$

$V(\textit{reject})$  is the prospect theory value of rejecting the gamble and keeping the status quo, and has a value of 0. When  $V(\textit{accept})$  is large and positive  $p(\textit{accept})$  will be near 1. When  $V(\textit{accept})$  is large and negative then  $p(\textit{accept})$  will be near 0. The *bias* parameter accounts for the general propensity to accept gambles regardless of their outcomes. This is often not included in prospect theory, but below we show that it is necessary. Setting  $bias = 0$  reduces the model to a more usual version of prospect theory with a stochastic Luce's choice rule. The sensitivity parameter  $\theta$  determines how sensitive the probability of accepting is to differences in prospect theory value, which is seen more clearly when Equation 4 is written in logistic form:

$$p(\textit{accept}) = \frac{1}{1 + e^{-(bias + \theta[V(\textit{accept}) - V(\textit{reject})])}} \quad (5)$$

Equations 3 and 4 can be rearranged to give the log-odds of accepting a gamble (Tom et al., 2007; Stewart, Reimers, & Harris, 2015).

$$\log \left[ \frac{p(\textit{accept})}{1 - p(\textit{accept})} \right] = bias + \theta w^{(1/2)} gain^\alpha - \theta w^{(1/2)} \lambda loss^\beta \quad (6)$$

When  $\alpha = \beta = 1$  Equation 6 is in the form of a logistic regression, where the choice of accepting is predicted as a function of *gain* and *loss*.

$$\log \left[ \frac{p(\textit{accept})}{1 - p(\textit{accept})} \right] = \phi_{intercept} + \phi_{gain} gain + \phi_{loss} loss \quad (7)$$

where

$$\phi_{intercept} = bias \tag{8}$$

$$\phi_{gain} = \theta w(1/2) \tag{9}$$

$$\phi_{loss} = -\lambda \theta w(1/2) \tag{10}$$

which means that

$$\lambda = -\frac{\phi_{loss}}{\phi_{gain}} \tag{11}$$

Thus is it convenient to estimate prospect theory parameters using a logistic regression, with  $\lambda$  set by Equation 11.

We can visualize how each parameter of the model plays a role in capturing the propensity to accept mixed gambles in a two dimensional gain-loss space. We demonstrate this with the responses taken from four randomly sampled participants in Experiment 1 of Walasek and Stewart (2015). In their version of the accept-reject task, participants saw 64 lotteries, which were constructed by combining losses and gains ranging from \$6.00 to \$20.00, in \$2.00 intervals. Figure 1 illustrates raw data from four participants, with green  $\circ$  symbols for accept responses and red  $\times$  for reject responses. Visibly, these participants behaved quite reasonably in that they tended to accept only those lotteries that offered a large gain and a small loss (the bottom right portion of gain-loss space).

Fitting the model in Equation 7 to each participant's responses yields parameter values that can be used to draw each person's indifference curve. When people are indifferent—that is, equally likely to accept or reject—then the log odds of accepting is zero. The equation for the indifference curve is given by setting  $\log \left[ \frac{p(\text{accept})}{1-p(\text{accept})} \right] = 0$  and rearranging Equation 6 with  $\alpha = \beta = 1$ .

$$loss = \frac{bias}{\lambda \theta w(1/2)} + \frac{1}{\lambda} gain \tag{12}$$



These indifference curves, drawn in Figure 1, clearly separate the area of acceptance and rejection for each participant. Two features of these lines are important to our argument: the slope, given by  $1/\lambda$ , and the intercept given by  $\frac{bias}{\lambda \theta w^{(1/2)}}$ . Thus the slope is determined by the loss aversion parameter  $\lambda$ . However, the intercept is determined by a combination of parameters:  $\lambda$ , the probability weighting for one half  $w^{(1/2)}$ , the sensitivity parameter  $\theta$ , and the overall tendency of accept gambles irrespective of the gain and loss on offer  $bias$ . When  $bias = 0$  this means the intercept is required to be zero, which is not the case for most participants. That is, the  $bias$  is necessary to account for people's individual differences in the overall propensity to accept or reject gambles independently of the gain and losses on offer.

To preempt the results of our simulations below, we are going to show that  $\lambda$  cannot be reliably estimated from accept-reject choice data. All that can be reliably estimated is the area under the indifference curve (*AUIC*) in Equation 12. This area corresponds to the fraction of gain-loss space in which people accept gambles. For a gain-loss space with uniformly distributed gains and losses, this corresponds quite simply to the fraction of gambles people accept. The intuition is that the indifference lines in Figure 1 can wiggle about their middles quite a bit, which changes the slope and the intercept a lot without changing the area under the line.

### Simulation 1

Our first simulation is a parameter recovery exercise, in which we generated simulated data from the model and then tried to recover the generating parameters by fitting the same model. We simulated responses in hypothetical experiments where participants must accept or reject 100 mixed gambles made up of every possible and unique combination of gains and losses ranging from \$2.00 to \$20.00, in \$2.00 increments. We used the parameters from the logistic regression form in Equation 7 to define our generating parameter grid. Our initial parameter values consisted of  $\phi_{gain}$  (ranging from .05 to 1 in .05

increments),  $\phi_{loss}$  (ranging from -1 to .05 in .05 increments) and  $\phi_{intercept}$  (ranging from -.2 to .2 in .2 increments). The entire parameter grid therefore consisted of all 8,400 combinations of these three parameters.

We converted the logistic-regression-form parameters into the corresponding prospect-theory parameters in Equation 6 using Equations 8–11.

For a given set of parameter values, we simulated 50 experimental sessions. This is analogous to a situation in which 50 individuals with the same underlying preferences take part in the same experiment.

For each choice we generated  $p(\textit{accept})$  with Equation 6. We then generated a response for each gamble—either accept or reject—as a Bernoulli sample based on  $p(\textit{accept})$  for that gamble. We excluded datasets for which the standard deviations of  $p(\textit{accept})$  over all trials was smaller than .2 to discard parameter values that make very similar predictions over all gambles in the choice set. As a result, we were left with 6,586 unique parameter combinations and 329,300 datasets (32,930,000 individual choices). The possible values of  $\lambda$  from our grid included 295 unique values ranging from .05 to 19 with a mean of 1.39 and SD of 1.46.

**Results.** We first removed 849 instances in which our model failed to converge, leaving us with 328,451 recovered parameter values. Figure 2 shows how well  $\lambda$ ,  $\textit{bias}$ ,  $\theta w(1/2)$ , and  $AUIC$  can be recovered for a random sample of 50,000 parameter values from our grid search. To summarise, while  $\lambda$ ,  $\textit{bias}$ ,  $\theta w(1/2)$  are poorly recovered,  $AUIC$  is nearly perfectly recovered.

Figure 2A shows that the recoverability of  $\lambda$  is poor. The recoverability worsens for higher levels of loss aversion, but even for the relatively lower values of  $\lambda < 2.5$ , the amount of heterogeneity is non-trivial in magnitude. Consider the case of true  $\lambda$  of 2.25, which is the often-used median value from (Tversky & Kahneman, 1992) and corresponds to pronounced loss aversion. Interpreting  $\lambda$  as the sole reason for rejecting mixed lotteries could lead us to believe that a person would require a gain to be approximately 2.25 higher

than a loss in order to be indifferent with respect to a given 50/50 gamble. However, the recovered  $\lambda$  is less than 1.99 30% of the time and more than 2.62 30% of the time. We therefore cannot know for sure what is the true  $\lambda$  of such a participant.

The ability to recover generating *bias* is even worse, as illustrated by Figure 2B. Similarly, in the case of  $\theta w^{(1/2)}$  (Panel C) the recovered values are very variable. The recoverability is particularly poor for high values of  $\theta w^{(1/2)}$  as indicated by the fanning out fits of the quantile regression. Appendix A further illustrates the large variance in the recovered parameters.

What does this variability in recovered values mean for loss averse behaviour? Answering this question requires us to distinguish between two possible interpretation of loss averse behaviour in the accept-reject task. One view is that loss aversion is simply given by the  $\lambda$  parameter of prospect theory, but another is that loss averse behaviour is simply the overall tendency to reject mixed gambles. To capture the overall tendency to reject mixed gambles we calculated the area under the indifference curve in gain-loss space (see Appendix B). For example, individuals whose indifference curves are plotted in the lower two panels of Figure 1 appear to differ considerably in their willingness to accept mixed lotteries in the accept-reject task, but this is not reflected in the slope of their indifference curves. Instead, the difference here is driven by the *bias* which contributes to the upward and downward shifts of this curve. If *bias* and  $\lambda$  (and also  $\theta w^{(1/2)}$ ) capture all there is about the propensity to accept mixed lotteries, we should expect to see very good recoverability of the *AUIC* scores.

Figure 2D illustrates recoverability of *AUIC*. Large *AUIC* scores mean that a larger portion of the gain-loss space is covered by the decisions to accept, and therefore high *AUIC* scores represent less loss averse behaviour. It is clear from the plot that *AUIC* is very well recovered. Of note is the small cloud in the lower right corner of the plot. These values emerge in rare cases when the binary accept-reject data generated from the model probabilities are particularly extreme. In such instances, the fitted model produces the

indifference curve that is outside the bounds of the gain-loss space.

If  $AUIC$  encompasses the global tendency to accept mixed gambles, then we should also find no correlations, indicative of parameter trade-offs, between  $AUIC$  and  $bias$ ,  $\theta w^{(1/2)}$  or  $\lambda$ . In contrast, we may expect a correlation between  $bias$  and  $\lambda$  if these two parameters truly trade-off in the model when they both try to capture decisions to accept and reject mixed lotteries. To demonstrate this, for each possible combination of the generating  $bias$ ,  $\theta w^{(1/2)}$  and  $\lambda$  parameters, we computed Spearman's correlation coefficients for recovered values of these parameters and  $AUIC$ . Figure 3 shows the distributions of correlation coefficients for all six possible pairings of parameters. There is a strong correlation between  $bias$  and  $\lambda$ . Recall that the elevation of the indifference curves in the two dimensional gain-loss space is given by the intercept  $\frac{bias}{\lambda w^{(1/2)}}$  and slope  $\frac{1}{\lambda}$ . The positive correlation coefficients therefore show that when the propensity to reject lotteries regardless of their outcomes is high, the asymmetric weighting of gains and losses is smaller. Conversely, we observe much weaker correlations between  $AUIC$  and  $\lambda$  and  $AUIC$  and  $bias$ . With respect to  $\theta w^{(1/2)}$  we observe weak correlations with the other two parameters as well as the  $AUIC$ .

Our simulation rather unambiguously shows that  $\lambda$  suffers from extremely poor recoverability when attempts are made to estimate it from the responses in the accept-reject task. This issue seems to be mainly driven by the parameter trade-off between  $\lambda$  and  $bias$  in the stochastic choice rule. A measure derived from both of these parameters, which corresponds to the area under a person's indifference curve in the gain-loss space, is very well-recovered. Since this measure captures the overall propensity to accept/reject mixed gambles, it could easily be replaced by simply calculating the proportion of "accept" decisions in the accept-reject task. We return to the topic of using  $AUIC$  in the accept-reject task in the Discussion.

## Simulation 2

In our second simulation, we determine whether the issue of poor recoverability of  $\lambda$  in the accept-reject task can be alleviated with a larger number of data points. That is, we attempt to establish what would be the minimum number of experimental trials that would produce a reliable measure of prospect theory's  $\lambda$ . We simulated data for 50 identical participants with true parameter values of the  $bias = 0.2203$ ,  $\phi_{gain} = 0.9673$  and  $\phi_{loss} = -1.2204$  — the median estimates from Walasek and Stewart (2015) Experiment 1A. For each simulated participant we simulated 64 responses to the 64 trials in the 20-20 Condition from Experiment 1A, where gains and losses were drawn from the set (\$6, \$8, \$10, \$12, \$14, \$16, \$18, and \$20). We repeated this process for a second simulated experiment in which the 64 trials were repeated twice (so 128 trials per simulated participant), a third experiment where the trials were repeated three times (so 192 trials per simulated participant), and so on, up to a somewhat impractical experiment where the 64 trials were repeated 60 times (so 3,840 trials per simulated participant).

**Results.** Density plots of recovered  $\lambda$ ,  $\theta w^{(1/2)}$ ,  $bias$  and  $AUIC$  are shown in four panels of Figure 4. For the single replication of the 64 trials, the distribution is a vertical slice from Figure 2. When there is one single replication of 64 trials, the interquartile range of the recovered  $\lambda$  values is 0.21. The distributions of recovered  $\theta w^{(1/2)}$  and  $bias$  are also very variable, with the interquartile range of 0.56 and 2.90, respectively. In contrast, the recoverability of  $AUIC$  is much better, with the interquartile range of 0.04 for the experiment with 64 trials.

As the number of repetitions is increased the standard deviations of the distributions of recovered parameter values reduce. But even for quite high numbers of repetitions,  $\lambda$ ,  $\theta w^{(1/2)}$ , and  $bias$  are still poorly recovered. For example, with two repetitions the interquartile range of the recovered  $\lambda$  values is 0.15. Even with 10 repetitions—640 trials per simulated participant in total—the interquartile range of the recovered  $\lambda$  values is 0.07.

## Discussion

One of the defining features of people’s preferences is their attitude towards risks involving gains and losses. It is therefore important that behavioural scientists have reliable and accurate tools for measuring an individual’s loss aversion. Here, we evaluated the popular accept-reject task for estimating the loss aversion parameter  $\lambda$  of prospect theory, in which participants must indicate whether they would accept or reject series of 50/50 gambles. We found that  $\lambda$  suffers from poor recoverability. This recoverability remains poor, in absolute terms, even with more than a thousand experimental trials. Our analysis shows very clearly that difficulties in estimating  $\lambda$  arises from parameter trade offs within the model. We also showed that a composite measure derived from both  $\lambda$ ,  $\theta w(1/2)$  and *bias*, which corresponds to portion of the gain-loss space in which people accept gambles (see Figure 1), is well recovered. This area is equivalent to the number of gambles accepted in a choice set with uniform distributions of gains and losses. Yet, while the *AUIC* measure is well recovered, it would not generalise to other choice sets with different ranges of gains and losses. As such, *AUIC* should not be regarded as a solution to the problem of estimating  $\lambda$  in the accept-reject task. At the level of the individual, any attempts to estimate prospect theory’s  $\lambda$  seem futile and can lead to serious mis-estimations.

By representing people’s decisions in the two-dimensional gain-loss space, we can also show that removing *bias* parameter from the exponentiated Luce’s choice rule would not be a good solution to the problem of estimating  $\lambda$  from accept-reject task data. Both the intercept and slope of the indifference curves plotted in Figure 1 are necessary to represent people’s attitudes towards mixed gambles. In other words, simplification of the model by eliminating the *bias* parameter would just result in poorer model fits. We checked this by fitting two models, one with *bias* = 0 and one with *bias* free to vary, for each participant in the experiments reported by Walasek and Stewart (2015). Comparing model fits using revealed that the *bias* = 0 constraint results in significantly poorer model fits for a considerable proportion of the participants (ranging from 15% to 48%, see Appendix ?? for

further details). Thus our argument is not that the model described in the Equation 6 is wrong, but rather that data provided by the accept-reject task are not suitable for recovering underlying attitudes towards gains and losses. The problem therefore emerges from the combination of information-poor binary choices and the complexity of the model necessary to capture people’s behaviour in full.

It is worth noting that *bias* may be particularly relevant in the context of the accept-reject task. Several authors pointed out that this type of task confounds loss aversion with the status quo bias (Yechiam & Hochman, 2013; Ert & Erev, 2013). Indeed, if participants are given an opportunity to choose between a \$0 for sure and a mixed gamble, they are much more likely to choose the gamble than they are to accept it in the accept-reject version of the task. So if the responses on the accept-reject task are in a large part driven by the status quo bias, then the issue of poor recoverability of  $\lambda$  will be worse in the case of this particular task.

How serious are these problems? In other words, could the variability of recovered values influence results of studies in which the accept-reject task is used to estimate  $\lambda$ ? The problem of poor recoverability will lead to mis-estimation on the individual level. This issue is particularly problematic for studies where individual estimates of  $\lambda$  are correlated with some other individual difference. Take as an example Tom et al. (2007), in which the authors found a strong positive correlation between participants’  $\lambda$  values estimated from the accept-reject task and neural loss aversion as measured by fMRI (specifically the differential sensitivity to gains and losses in dopaminergic regions of the brain). Using the results of our grid search in Simulation 1, we can demonstrate that such correlations are unreliable when the accept-reject task is used to estimate  $\lambda$ . We begin by simultaneously sampling a set of “true”  $\lambda$  values from a normal distribution with a mean of 2.25 and standard deviation of 1.22 (taken from Lorains et al. (2014) as Tom et al. (2007) did not report the mean and standard deviation) and a set of “true” neural loss aversion values that correlate  $r = 0.65$  with the “true”  $\lambda$ s. Actually Tom et al. (2007) report an even

higher correlation of  $r = 0.85$ , but our finding is robust across different values of  $r$ . For each “true”  $\lambda$  we drew another corresponding “recovered”  $\lambda$  from the distribution in Figure 2. We then calculated the correlation between the “recovered”  $\lambda$  values and the “true” neural loss aversion. We repeated this 1,000 times for sample sizes of 15, 20, 25, and 35 participants. The distribution of 1,000 correlations is plotted in Figure 5 for each sample size. The recovered correlations were considerably lower than the starting value of 0.65 (blue dotted line). The underestimation is large, with the correlations ranging from  $-.73$  to  $.77$  (median =  $.10$ ). This result should be worrying to researchers who wish to use the accept-reject task to quantify individual differences in loss aversion.

We think that the issue of poor parameter recoverability is less bad when making estimates about a group of participants rather than individual participants. Recall that the relationship between the standard deviation of a sample and the standard error of the mean—the standard error is the standard deviation divided by the square root of the sample size. For example, while the standard deviation of the recovered  $\lambda$ s in Simulation 2, with one replication of 64 trials for each of the 50 simulated participants is 0.15 around a true mean of  $\lambda = 1.26$ . The standard error is  $0.15/\sqrt{50} = 0.02$ . Thus the estimate of the mean of the population from which the sample of participants was drawn will be well estimated, given a sufficient number of participants, even though the estimate for each individual participant is poor: While the accept-reject task provides a poor estimate of any given individual’s value of  $\lambda$ , the accept-reject task can offer a adequately precise estimation of the population mean.

In Simulation 2, we showed that merely increasing the number of trials does not solve the problem of poor recoverability of  $\lambda$  for individuals. However, another possibility would be to change the number of trials more strategically, increasing the number of lotteries in the regions that are most diagnostic for determining the slope of one’s indifference curve (e.g., at the edges of the gain-loss space). Even if an experimenter had a good idea about the reasonable set of lotteries for recovering  $\lambda$ , manipulating sets of



lotteries in the accept-reject task can have serious consequences. Stewart, Canic, and Mullett (2020) and André and de Langhe (2019) both show that the sets of questions used to elicit prospect theory parameters biases the estimate of those parameters. Stewart et al. (2020) show that the correlation between prospect theory parameters estimated across a random partitioning of a choice set is disturbingly low: estimated parameters do not even generalise to a new set of choices drawn from the same master set. André and de Langhe (2019) show that allowing the range of gains and losses to vary biases the estimate of  $\lambda$  in the accept-reject task. Thus if one is to have a measure of  $\lambda$  which is comparable across participants—as is required for correlating  $\lambda$  with other individual differences—one must use the same set of choices over all individuals. This requirement thus rules out having the choices generated for each participant from an adaptive procedure.

Our finding that the accept-reject task is not suitable for recovering prospect theory's  $\lambda$  involved using Luce's choice rule as an error model to allow prospect theory to make stochastic predictions. This extension is often used and, for the accept-reject task, is equivalent to a logistic regression. We believe that the poor recoverability of  $\lambda$  would hold for alternative error models. Here we consider two prominent alternative approaches (see Loomes, 2005, for discussion). The first approach is to assume that people select the alternative (i.e., accept or reject) with the higher prospect theory value, but that their hand trembles and sometimes selects the other response. To see how this trembling-hand model is similarly affected, consider the probability of accepting a gamble as a surface over gain-loss space, rising as the gain grows and the loss falls. Under the Luce choice model (Equation 5), this is a smooth surface with a logistic shaped cross section. This trembling hand model effectively replaces the smoothly curving surface with a step function, stepping at the point where people are indifferent. But the same issue remains. Loss aversion is about identifying the orientation of the indifference curve in gain-loss space. All switching to the trembling-hand form has done is change the shape of the probability surface. In particular, because the trembling hand model has the same probability of tremble, no

matter what the gain or loss, the model equalises relative importance of the responses far from the indifference line in setting the orientation of the line. This does not make it easier to estimate the orientation of the indifference line. The same is true of a second prominent approach. In the Fechner model the difference in prospect theory value between accept and reject is subject to Gaussian noise and the choice is determined by the sign of the noisy difference. Thus the logistic form of the probability surface from the Luce model is replaced with a cumulative Gaussian form. Because of the very similar estimates obtained from logistic regression and probit regression (where the logistic form is swapped for a cumulative Gaussian), the same issue with the poor recoverability of  $\lambda$  will apply. Again, estimating  $\lambda$  is about estimating the orientation of the indifference line in gain-loss space, and subtly changing the curvature of the surface does not fix this problem. In short, the non-recoverability of  $\lambda$  will apply not just to the Luce's choice error model, but also to the trembling hand and Fechner error models.

The accept-reject task, as generally used, is clearly not appropriate for estimating prospect theory's  $\lambda$  at the level of an individual. A research who uses this task would be better off with simply counting the number of accept decisions. While the present paper does not offer a simple fix to the recoverability issue, it adds to the growing body of evidence demonstrating serious limitations of risky choice tasks for estimating  $\lambda$  under parametric assumptions.

## References

- Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty*, *36*, 245–266. doi: 10.1007/s11166-008-9039-8
- André, Q., & de Langhe, B. (2019). *No evidence for loss aversion disappearance and reversal in Walasek and Stewart (2015)*. Manuscript submitted for publication.
- Canessa, N., Crespi, C., Motterlini, M., Baud-Bovy, G., Chierchia, G., Pantaleo, G., . . . Cappa, S. F. (2013). The functional and structural neural basis of individual differences in loss aversion. *Journal of Neuroscience*, *33*, 14307–14317. doi: 10.1523/JNEUROSCI.0497-13.2013
- Chib, V. S., De Martino, B., Shimojo, S., & O'Doherty, J. P. (2012). Neural mechanisms underlying paradoxical performance for monetary incentives are driven by loss aversion. *Neuron*, *74*, 582–594. doi: 10.1016/j.neuron.2012.02.038
- Chumbley, J. R., Krajbich, I., Engelmann, J. B., Russell, E., Van Uum, S., Koren, G., & Fehr, E. (2014). Endogenous cortisol predicts decreased loss aversion in young men. *Psychological Science*, *25*, 2102–2105. doi: 10.1177/0956797614546555
- Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making*, *8*(3), 214–235. Retrieved from <http://journal.sjdm.org/12/12712/jdm12712.html>
- Frydman, C., Barberis, N., Camerer, C., Bossaerts, P., & Rangel, A. (2014). Using neural data to test a theory of investor behavior: An application to realization utility. *Journal of Finance*, *69*, 907–946. doi: 10.1111/jofi.12126
- Gal, D., & Rucker, D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, *28*. doi: 10.1002/jcpy.1047
- Gelskov, S. V., Henningson, S., Madsen, K. H., Siebner, H. R., & Ramsøy, T. Z. (2015). Amygdala signals subjective appetitiveness and aversiveness of mixed gambles. *Cortex*, *66*, 81–90. doi: 10.1016/j.cortex.2015.02.016

- Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*, 21–32. doi: 10.1016/j.cognition.2011.12.002
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, *5*, 193–206.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291. doi: 10.2307/1914185
- Lazzaro, S. C., Rutledge, R. B., Burghart, D. R., & Glimcher, P. W. (2016). The impact of menstrual cycle phase on economic choice and rationality. *PLoS ONE*, *11*, 1–15. doi: 10.1371/journal.pone.0144080
- Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, *8*, 301–323. doi: 10.1007/s10683-005-5372-9
- Lorains, F. K., Dowling, N. A., Enticott, P. G., Bradshaw, J. L., Trueblood, J. S., & Stout, J. C. (2014). Strategic and non-strategic problem gamblers differ on decision-making under risk and ambiguity. *Addiction*, *109*, 1128–1137. doi: 10.1111/add.12494
- Markett, S., Heeren, G., Montag, C., Weber, B., & Reuter, M. (2016). Loss aversion is associated with bilateral insula volume. A voxel based morphometry study. *Neuroscience Letters*, *619*, 172–176. doi: 10.1016/j.neulet.2016.03.029
- Pachur, T., & Kellen, D. (2013). *Modeling gain-loss asymmetries in risky choice: The critical role of probability weighting*. Retrieved from <https://mindmodeling.org/cogsci2013/papers/0571/paper0571.pdf>
- Pighin, S., Bonini, N., Savadori, L., Hadjichristidis, C., & Schena, F. (2014). Loss aversion and hypoxia: less loss aversion in oxygen-depleted environment. *Stress: the International Journal On the Biology of Stress*, *17*, 204–210. doi: 10.3109/10253890.2014.891103
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of*

- Experimental Psychology: Learning, Memory, and Cognition*, 34, 1446–1465. doi: 10.1037/a0013646
- Rieskamp, J., Busemeyer, J. R., & Mellers, B. (2006). Extending the bounds of rationality: A review of research on preferential choice. *Journal of Economic Literature*, 44, 631–661. doi: 10.1257/jel.44.3.631
- Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, 22, 391–407. doi: 10.3758/s13423-014-0684-4
- Schulreich, S., Heekeren, H. R., & Gerhardt, H. (2016). Incidental fear cues increase monetary loss aversion. *Emotion*, 16(3), 402–412. doi: 10.1037/emo0000124
- Sokol-Hessner, P., Camerer, C. F., & Phelps, E. A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social Cognitive and Affective Neuroscience*, 8, 341–350. doi: 10.1093/scan/nss002
- Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 5035–5040. doi: 10.1073/pnas.0806761106
- Sokol-Hessner, P., Lackovic, S. F., Tobe, R. H., Camerer, C. F., Leventhal, B. L., & Phelps, E. A. (2015). Determinants of propranolol's selective effect on loss aversion. *Psychological Science*, 26, 1123–1130. doi: 10.1177/0956797615582026
- Stewart, N., Canic, E., & Mullett, T. L. (2020). *On the futility of estimating utility functions: Why the parameter we measure are wrong, and why they do not generalize*. Manuscript submitted for publication. Retrieved from <https://psyarxiv.com/qt69m> doi: 10.31234/osf.io/qt69m
- 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. doi: 10.1287/mnsc.2013.1853

- Stewart, N., Scheibehenne, B., & Pachur, T. (2018). *Psychological parameters have units: A bug fix for stochastic prospect theory and other decision models*. Manuscript submitted for publication. Retrieved from <https://psyarxiv.com/qvgcd> doi: 10.31234/osf.io/qvgcd
- Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*, 101–130. doi: 10.1007/s11166-006-8289-6
- Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007, January 26). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323. doi: 10.1007/BF00122574
- Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press.
- 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. doi: 10.1037/xge0000039
- Yechiam, E. (2018). Acceptable losses: The debatable origins of loss aversion. *Psychological Research*, *83*, 1327–1339. doi: 10.1007/s00426-018-1013-8
- 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. doi: 10.1037/a0029383

Table 1

*Summary of the studies using the accept-reject task to estimate  $\lambda$*

Author(s)	Year	N Participants	N Trials	$\lambda$
Barkley-Levenson et al.	2013	16	144	NA
Canessa et al.	2013	56	104	NA
Chib et al.	2012	32	512	m = 2.09
Chumbley et al.	2014	43	36	NA
Frydman et al.	2011	90	140	$\mu = 1.52$
Gelskov et al.	2015	16	64	$\mu = 2.16, m = 2.08$
Lazzaro et al.	2016	39	100	$\mu = 1.46-1.78$
Lorains et al.	2014	38	49	$\mu = 2.25$
Markett et al.	2016	46	256	$\mu = 1.89, m = 1.66$
Pighin et al.	2014	26	256	$\mu = 2.37$
Schulreich et al.	2016	29	200	NA
Sokol-Hessner et al.	2013	47	150	$\mu = 1.31, 1.15$
Sokol-Hessner et al.	2009	60	140	$\mu = 1.62, 1.41$
Sokol-Hessner et al.	2015	22	180	$\mu = 1.38$
Tom et al.	2007	16	256	$\mu = 1.93$

*Notes.* Aggregate  $\lambda$  values are reported for control groups only, unless it was not clear which of the available groups acted as a control (in such cases estimates from both groups are included).  $\lambda$  values are means, except for Chib et al. which is a median. Table provides aggregate estimates of  $\lambda$ , when authors reported either, both, or neither (NA).

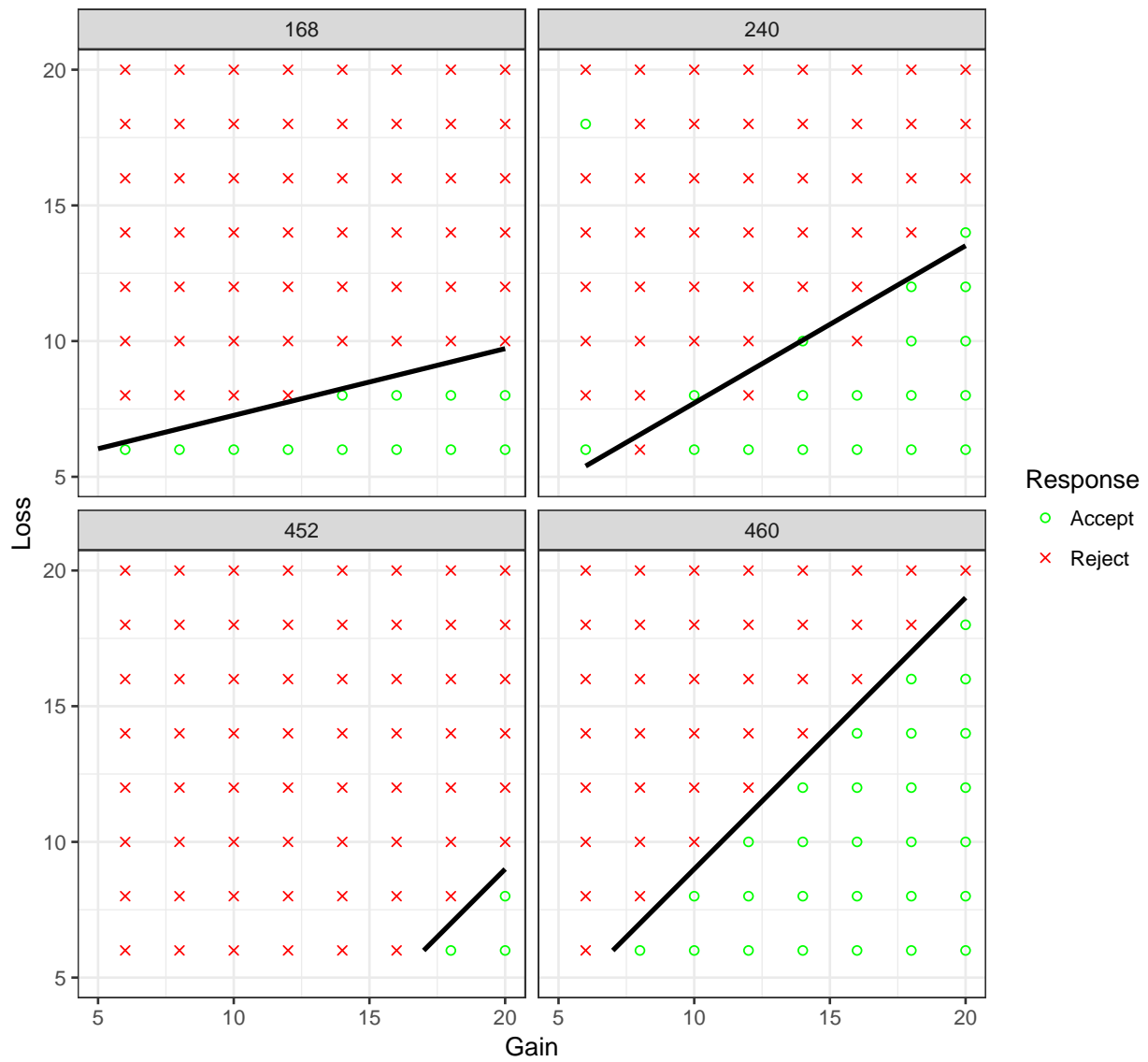
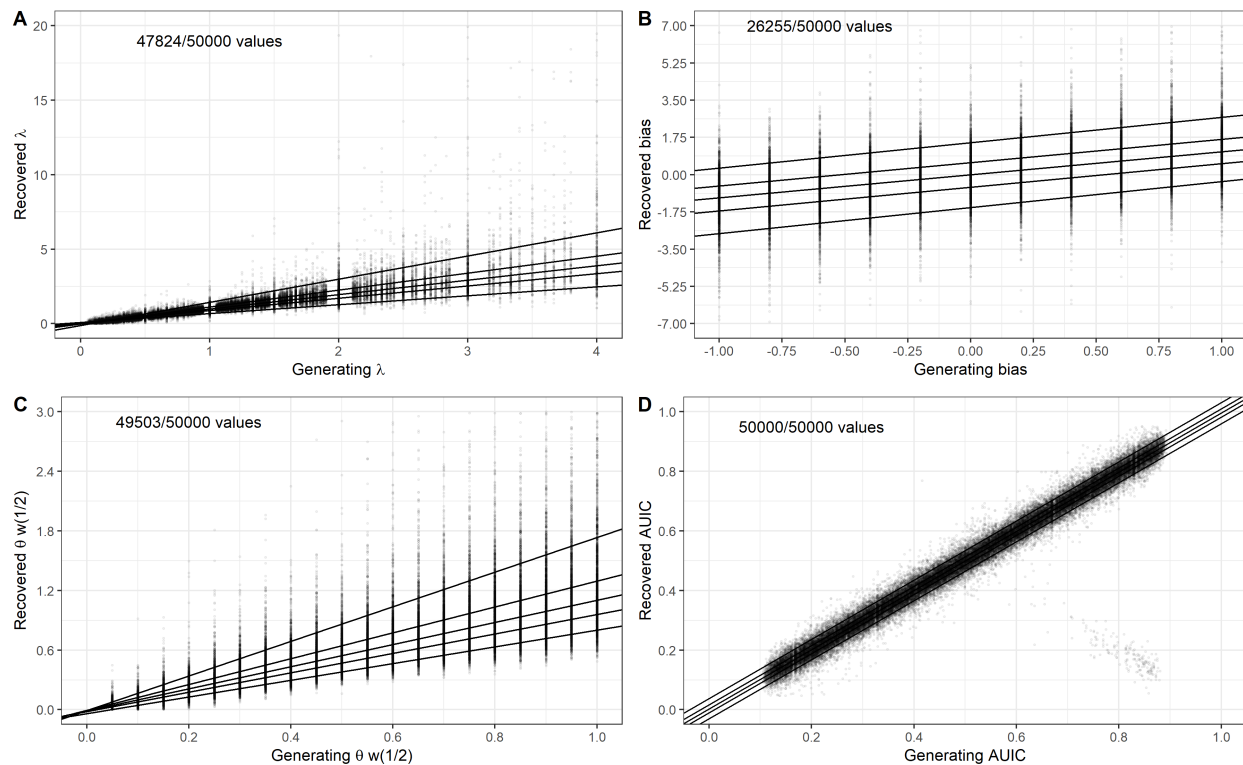


Figure 1. Example responses and fitted indifference curves of four participants from Walasek and Stewart (2015).





*Figure 2.* Scatterplots of recovered parameter values plotted against generating parameters. Each plot is based on the same random sample of 50,000 data points from the parameter recovery. Panel A: Recovered  $\lambda$  against generating  $\lambda$ . Panel B: Recovered *bias* against generating *bias*. Panel C: Recovered  $\theta w(1/2)$  against generating  $\theta w(1/2)$ . Panel C: Recovered *AUIC* against generating *AUIC*. In Panels A, B and C, the y-axis scale has a larger range than the x-axis so that more of the recovered values can be displayed. Nonetheless, some parameter values are not shown; the number of values shown is displayed inside the panel.

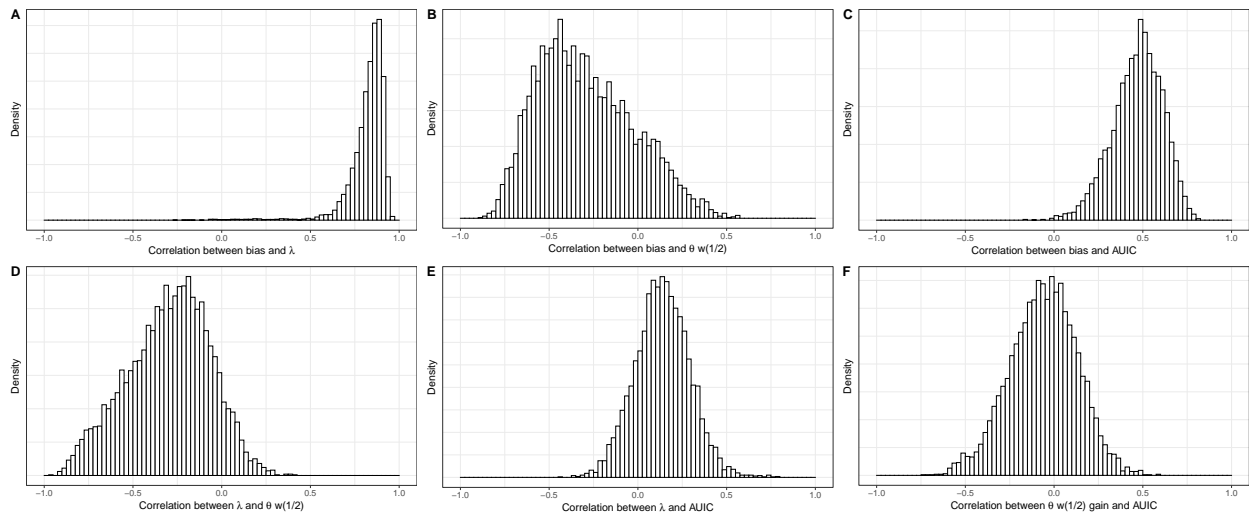
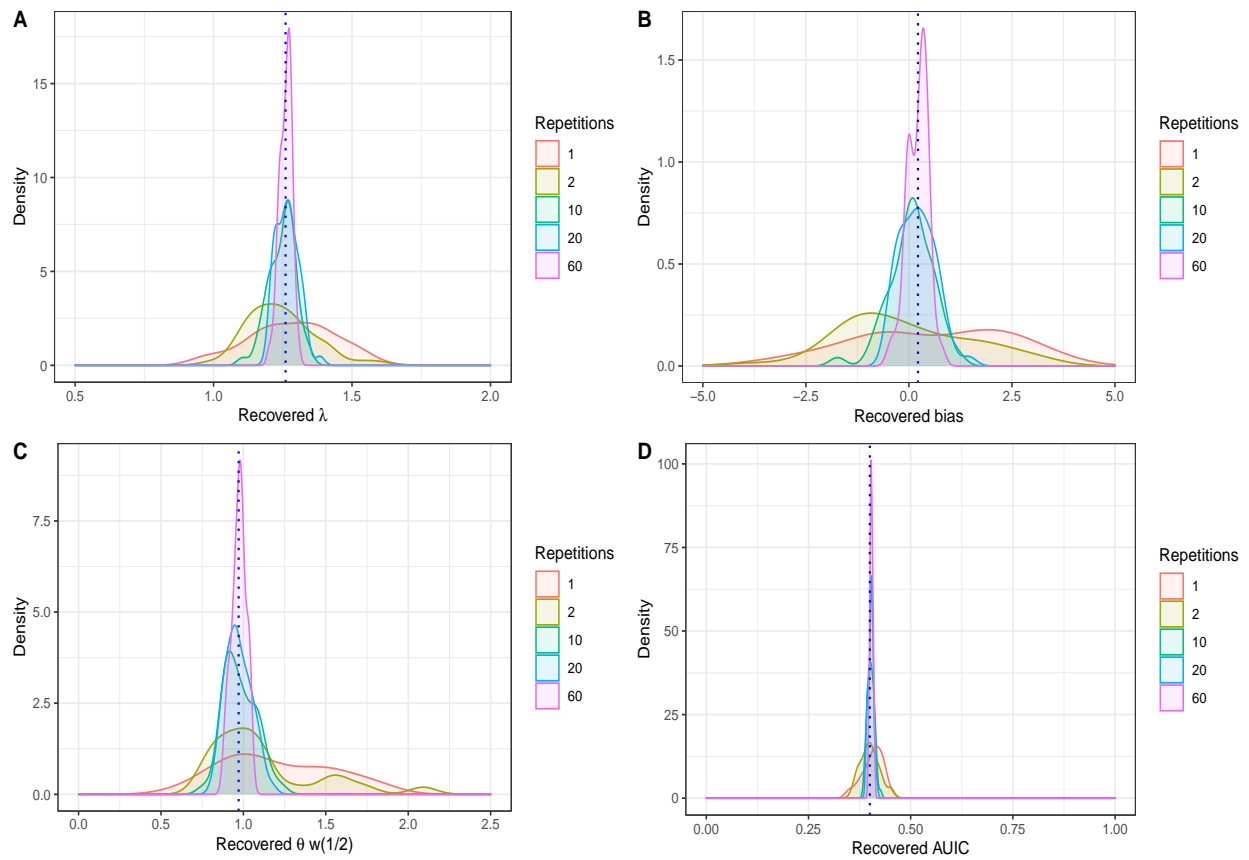


Figure 3. Distributions of Spearman's correlation coefficients for pairs of parameter values:  $\lambda$ , bias,  $\theta w(1/2)$ , and AUIC.



*Figure 4.* Distributions of recovered parameters. Panel A:  $\lambda$ , Panel B: *bias*, Panel C:  $\theta w(1/2)$ , and Panel D: *AUIC*, for selected numbers of repetitions. Dashed vertical lines show the true parameter value, identical for all simulated participants. Distributions are the distributions of the recovered parameter value over 50 identical simulated participants.

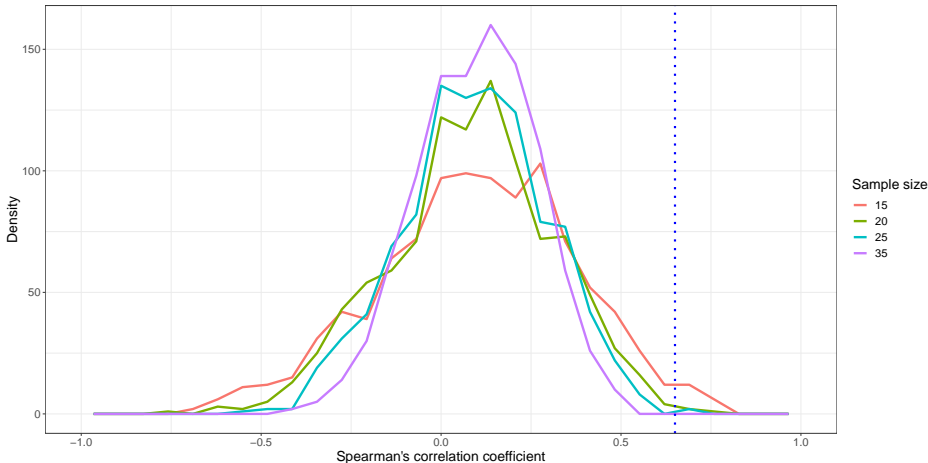


Figure 5. Density plots of recovered Spearman's correlation coefficients for varying sample sizes. The dotted blue line represents correlations of the initial dataset.

## Appendix A

## Alternative Parameter Recovery Plot

Figure A1 replots data from Figure 2 to highlight the median, 25th, and 75th percentiles of recovered parameters.

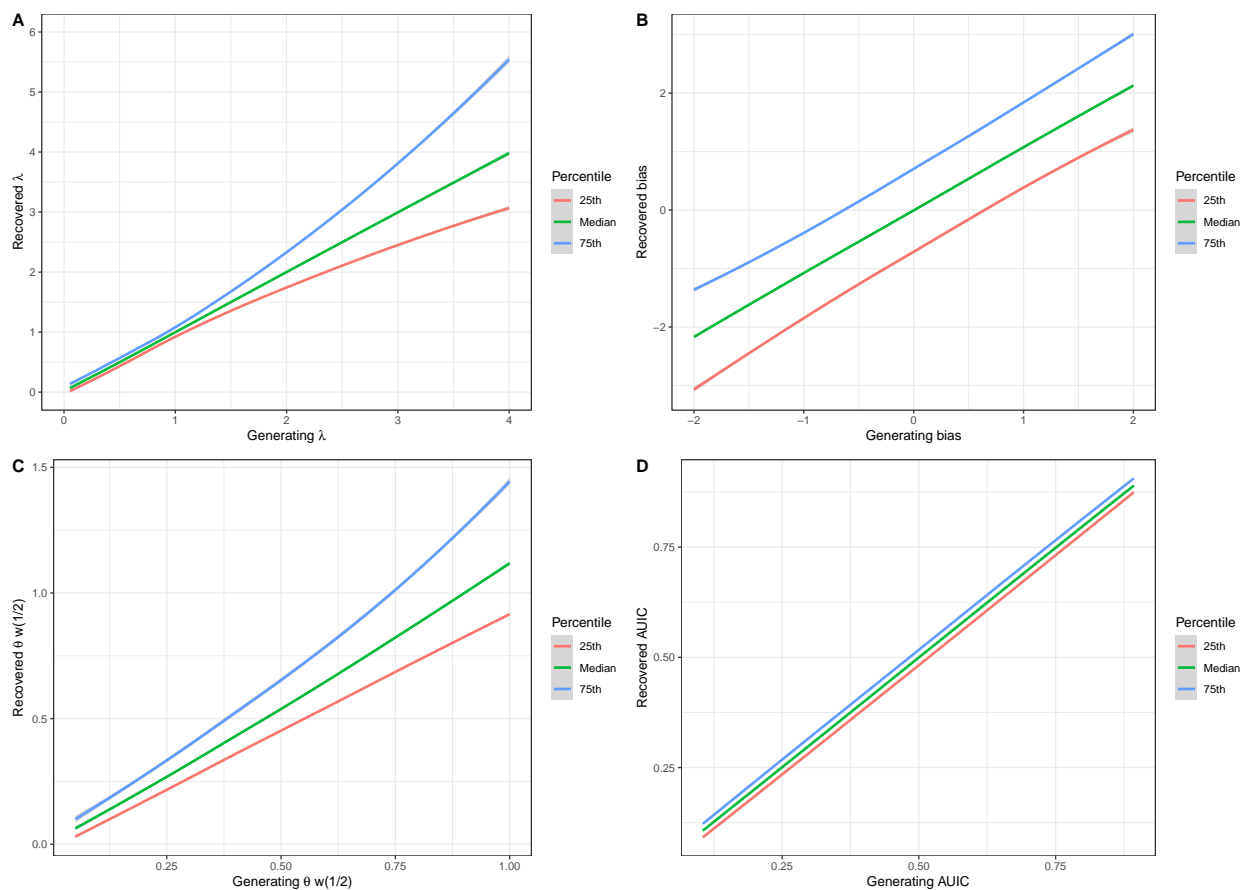


Figure A1. Median, 25th, and 75th percentile recovered parameter values plotted against generating parameters Panel A: Recovered  $\lambda$  against generating  $\lambda$ . Panel B. Recovered bias against generating bias. Panel C: Recovered  $\theta w(1/2)$  against generating  $\theta w(1/2)$ . Panel C: Recovered AUIC against generating AUIC.

## Appendix B

## Calculating Area Under the Curve

Consider a gain-loss space square with gains running horizontally from 0 to  $G$  and losses vertically from 0 to  $L$ . The indifference line, along which the decision maker is equally likely to accept or reject the gamble is given by Equation 12.

The intercept  $\beta_0 = \frac{\text{bias}}{\lambda \theta w^{(1/2)}}$  determines whether the indifference line enters the gain-loss square along the bottom or on the left. Together the intercept and the slope  $\beta_1 = 1/\lambda$  determine whether the line exits the gain-loss square along the top or on the right. (Avoid confusing intercept  $\beta_0$  and slope  $\beta_1$  with risk aversion in the loss domain  $\beta$  from Equation 1.) Thus there are four cases in Table B1 depending on the slope and the intercept for the area under the curve. The expressions are straightforward calculations resulting from adding and subtracting the areas of rectangles and triangles in the gain-loss square.

Table B1

*Closed-form expressions for area under the curve*

Entry	Exit	AUIC Expression
Left, $\beta_0 \geq 0$	Top, $\beta_0 + \beta_1 G \geq L$	$LG - (L - \beta_0)^2 / (2\beta_1)$
Left, $\beta_0 \geq 0$	Right, $\beta_0 + \beta_1 G < L$	$G(2\beta_0 + \beta_1 G) / 2$
Bottom, $\beta_0 < 0$	Top, $\beta_0 + \beta_1 G \geq L$	$L(2\beta_0 + 2\beta_1 G - L) / (2\beta_1)$
Bottom, $\beta_0 < 0$	Right, $\beta_0 + \beta_1 G < L$	$(\beta_0 + \beta_1 G)^2 / (2\beta_1)$

## Appendix C

Model fit comparison: with and without *bias*

We fitted two versions of the logistic regression to individual-level choice data from Walasek and Stewart (2015). We then performed a Likelihood Ratio Test to compare performance of the two models to capture each individual's responses. Table C1 summarizes the proportions of participants for whom there was a significant difference in the model performance, such that the model with *bias* offered a better model fit than a model without *bias*.

Table C1

*Proportion of participants best fitted by a model with a bias parameter*

---

Experiment	Proportion
1A	0.15
1B	0.14
2	0.20
3	0.48

---