Shift of probability weighting by joint and separate evaluations: Analyses of cognitive processes based on behavioral experiment and cognitive modeling

Yutaro Onuki (onuki-yutaro32@g.ecc.u-tokyo.ac.jp)
Graduate School of Arts and Sciences, The University of Tokyo
3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

Hidehito Honda (hitohonda.02@gmail.com)
Department of Psychology, Yasuda Women's University
6-13-1, Yasuhigashi, Asaminami-ku, Hiroshima 731-0153, Japan.

Toshihiko Matsuka (matsuka.toshihiko@gmail.com)
Department of Cognitive and Information Science, Chiba University
1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan.

Kazuhiro Ueda (ueda@gregorio.c.u-tokyo.ac.jp)
Graduate School of Arts and Sciences, The University of Tokyo
3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

Abstract

We examined whether probability weighting in decisions made under risk changed depending on the difference in evaluation methods. In particular, we focused on two methods, joint evaluation (JE) and separate evaluation (SE). We conducted a behavioral experiment and found that participants put more probability weight on small probability when using the SE method than when using JE, and that for large probabilities, the inverse was observed (i.e., participants put more weight in JE). We analyzed these results using a cognitive model and found that participants' subjective value of money does not change owing to differences in evaluation methods. However, beliefs concerning uncertain events shifted depending on evaluation methods, which led to the differences in probability weight. In this paper, we also discuss psychological mechanisms that produce different judgments or evaluations between SE and JE.

Keywords: probability weight; separate evaluation; joint evaluation; computer simulation; cognitive model of decision making

Introduction

It is well known that judgments change greatly depending on the difference in the evaluation methods. In the present study, we focused on one of the most studied topics, the difference between separate evaluation (hereafter, SE) and joint evaluation (hereafter, JE; Hsee, 1996; Hsee, Loewenstein, Blount, & Bazerman, 1999). Hsee (1996) showed that preference reversals by SE and JE occur in several contexts. Imagine people evaluating the worth of the following two dictionaries:

Dictionary A: Number of entries, 10,000
Dictionary B: Number of entries, 20,000 (cover is broken)

When they evaluate dictionaries A and B at the same time (i.e., JE), they may easily spot that there is a difference in the number of entries, and they may be attracted by the number of entries in Dictionary B. Thus, they may evaluate Dictionary B as having a higher price than Dictionary A. However, if people evaluate these dictionaries separately (i.e., SE), they may not notice the difference in the total number of entries (they may feel that either is enough), but they may mind the broken cover of Dictionary B. Thus, they may value Dictionary A as having a higher price than Dictionary B.

We predicted that shifts in evaluations by JE and SE might occur in the evaluation of probabilistic information. Previous studies on decisions under risk have shown that people put unique weights (i.e., non-linear weight) on probabilistic information in making decisions (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). For example, in decisions under risk, although people tend to be highly sensitive to differences in the end point (e.g., the difference between 0% and 10%, or differences between 90% and 100%), they tend to be less sensitive to differences in the middle degree (e.g., the difference between 30% and 40%). This finding suggests that sensitivity to differences is not constant. Recent studies have also showed that probability weighting is constructed through experimental procedures. In particular, different sets of probabilistic values presented in experimental tasks induce different probability weighting (e.g., Stewart, Reimers, & Harris, 2014; Walasek & Stewart, 2015).

Based on these previous findings, we predicted that differences in evaluation between JE and SE would change the probability weighting. If so, then what differences will be generated between JE and SE? In evaluating a certain probability value, people may refer to their probabilistic beliefs. For example, in evaluating 30% in a probabilistic event, people may refer to their probabilistic beliefs (i.e., how likely is the event to occur) and compare 30% with that belief. If they believe that the event usually occurs with high probability,

Money (Yen)	Prefer Sure Thing	Prefer Gamble
9500	✓	
9000	✓	
8500	✓	
8000	✓	
7500	✓	
7000	✓	
6500	✓	
6000	✓	
5500	✓	
5000	✓	
4500	✓	
4000	✓	
3500	✓	
3000	✓	
2500	✓	
2000	✓	
1500		✓
1000		✓
500		✓

Figure 1. A stimulus for measuring CE. The checks indicate the participant's selection.

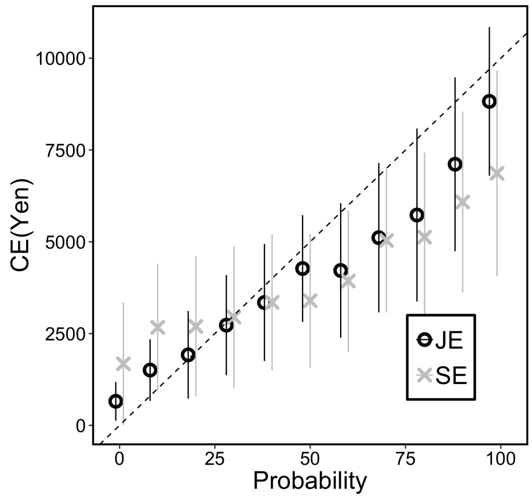

Figure 2. Mean CEs for 11 probabilities in the two groups. Error bars show standard deviation.

Table 1. Results of statistical analyses about the difference in CE between the two groups.

Probability	*t*-test		Effect size (*d*)
1	t (99) = 4.00	p = .001	0.80
10	t (106) = 4.23	p < .001	0.82
20	t (109) = 2.45	p = .174	0.47
30	t (118) = 0.68	p = .999	0.13
40	t (95) = 0.01	p = .999	0.00
50	t (93) = 2.58	p = .125	0.53
60	t (111) = 0.80	p = .999	0.15
70	t (101) = 0.19	p = .999	0.04
80	t (98) = 1.28	p = .999	0.26
90	t (99) = 2.13	p = .392	0.42
99	t (101) = 3.99	p = .001	0.79

Note. *p*-value was adjusted with *Bonferroni's* method.

they may judge 30% as "not enough." In contrast, if they believe that the event usually occurs with low probability, they may judge 30% as "enough." Then, what is the nature of people's belief about probabilistic events? Stewart, Chater, and Brown (2006) showed that when people communicate probabilistic information using verbal expressions such as "likely" or "impossible," they tend to use highly extreme expressions such as "never" (representing 0%) or "always" (representing 100%). This finding suggests that people tend to easily imagine event occurrences or non-occurrences. In other words, people may refer to "black and white" probabilistic beliefs when evaluating probability.

We predicted that this would be true in evaluations using the SE method, but that it may not be true in evaluations using the JE method. In JE, people are presented with some probabilistic values at the same time, and they can compare these values. Thus, people may refer to probabilistic information in a continuous way. To the best of our knowledge, no previous studies have examined the above issue. In the present study, using a behavioral experiment and cognitive modeling, we examined whether probability weighting would shift depending on differences in the evaluation method between SE and JE. In the following section, we report the results of our behavioral experiment. We then report our analyses based on cognitive modeling.

Behavioral experiment

We examined whether probability weighting would shift due to using different evaluation methods—specifically, JE or SE—in a gambling task.

Method

Participants. We recruited 682 students as participants.

Task, stimulus, and procedure. We followed the method in Gonzalez and Wu (1999) to conduct the following task: Participants were asked to make a choice between a gamble that gets 10,000 yen (around $100) with certain probability p or sure gain. Figure 1 shows an example of the task. For example, participants choose one of two options: 100% chance of winning 5,000yen or a 30% chance of winning 10,000 yen. When they choose to a sure option, the monetary value of the option decreased: "you can get 4,500 yen." Amounts of sure gain ranged from 9,500 yen to 500 yen. As seen in Figure 1, the choice should change from a sure option to a gamble, and in the change point, we can assume that there is an amount of money to which a person is indifferent about getting a sure gain or playing the gamble (i.e., certainty equivalent, hereafter, CE). We assumed that CE was the median of the change point (in Figure 1, CE was assumed to be 1,750 yen). For the probability of the gambles, we set 11 values: 1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 99%.

In the JE group (n = 47), participants were asked to answer the choices for the 11 probabilities. At first, they were instructed to answer the choices for 11 probabilities and then check their choices for each while answering the questions. In the SE group (n = 635), they were presented with one of

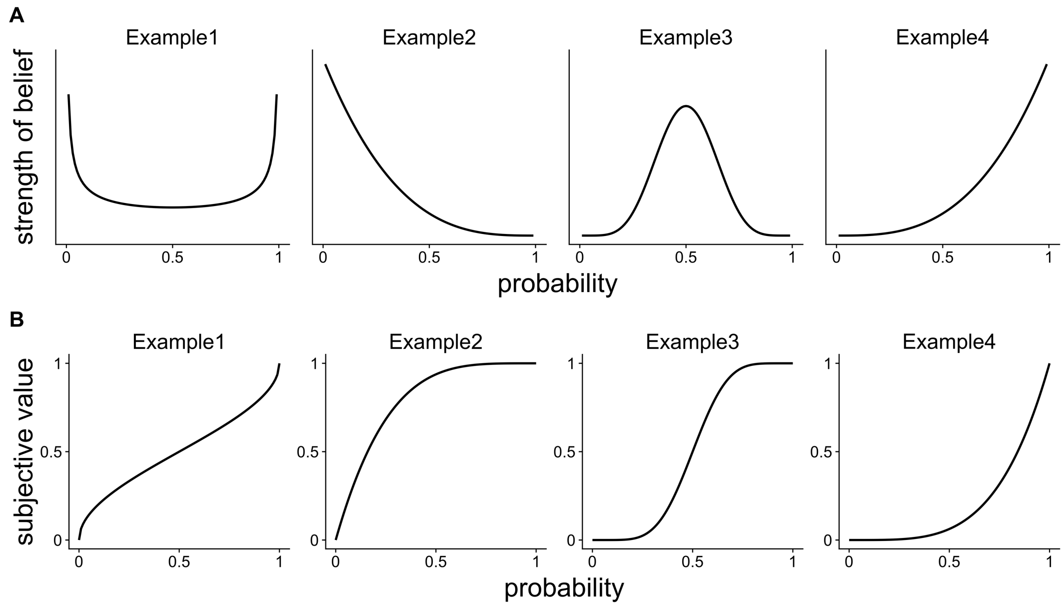

Figure 3. Summaries of DbBS. (a) Probabilistic belief regarding an uncertain event. (b) Subjective value in DbBS. This is represented with the cumulative distribution function (CDF) of the beta distribution.

the 11 probability gambles and answered the choices for the probability.

Results

Figure 2 shows mean CEs for 11 probabilities in the two groups. We found that the CEs differed between the two groups. In particular, in the low-probability range (1–20), the CE was higher for the SE group than in the JE group, suggesting that participants in the SE group applied more probability weight than those in the JE group. However, this trend reversed in the high-probability range (80–99), suggesting that participants in the JE group applied more probability weight than those in the SE group (as to the statistical analyses of CEs, see Table 1).

Taken together, the difference in evaluations between the JE and SE groups induced different probability weighting. In the following sections, we report the analyses of cognitive processes using a cognitive model.

Analyses of cognitive processes based on the cognitive model

Cognitive model of decision making: Decision by Belief Sampling (DbBS)

In this section, we introduce the decision model, called the *decision by belief-sampling model* (hereafter, DbBS; Honda, Matsuka, & Ueda, 2017). This model was proposed based on the *decision by sampling* model (*DbS*; Stewart, Chater, & Brown, 2006; Stewart, 2009). In the DbS model, subjective attribute values are constructed by a series of binary, ordinal comparisons to a sample of attribute values that reflect the immediate decision context and real-world distribution. The subjective value for a target is calculated as follows:

$$r = \frac{R-1}{N-1} \tag{1}$$

where r ($0 \leq r \leq 1$) denotes the subjective value for a target, and R denotes the rank of the target within the decision sample of N items. In this model, if the decision sample differs, r varies in the relationship between R and the decision sample. For example, imagine the subjective value for 60%. When decision samples are 10%, 20%, 30%, 30%, and 70%, the subjective value is $r = (5\text{-}1)/(6\text{-}1) = 0.8$. In contrast, in decision samples of 20%, 30%, 70%, 80%, and 90%, the subjective value is $r = (3\text{-}1)/(6\text{-}1) = 0.4$. That is, even when the target has the same attribute value, the subjective value varies depending on the decision samples. Previous studies have shown that this model can explain evaluations that vary depending on the samples (e.g., Stewart, Chater, Stott, & Reimers, 2003; Stewart, Reimers, & Harris, 2014).

DbBS is a model representing the subjective evaluation of probability. DbBS has two assumptions. First, the decision maker (DM) refers to the probabilistic belief samples in making decisions, and these samples represent the DM's probabilistic belief of an event's occurrence. For example, imagine the probable success rates of medical procedures for a serious disease and for appendicitis, respectively. Generally, people believe that the probability of success in treating a serious disease is low compared to the probable success of treating something simple, like appendicitis (Honda & Matsuka, 2014). We assume that the DMs refer to belief samples according to their probabilistic beliefs. We represent these beliefs using beta distributions (see the four examples of DMs' subjective beliefs in Figure 3[a]). Example 1 represents the belief such that an event will occur or not (people refer to event occurrence and nonoccurrence). Likewise, in Examples 2 and 4, the DMs have the belief such that the event will happen with a low or high probability. Example 3 represents the belief that an event has a 50% chance of occurring. Thus, beta distributions can represent extensive kinds of beliefs about uncertain events. As a second assumption, a subjective value for a target is constructed by the comparison between the target value and the belief samples. Figure 3(b) shows subjective values calculated by the DbBS model. Given that beta distributions represent beliefs about uncertain events, subjective values correspond to values in the cumulative distribution functions (CDF) of beta distributions.

Using DbBS, we estimated the beliefs participants had in answering the gambling task in the behavioral experiment. In particular, we focused on the difference in beliefs produced between participants in the JE and SE groups.

Parameter estimation

In the gambling task of the behavioral experiment, when CE is y yen for the gamble that can win 10,000 yen with probability p, we assumed that the following relation:

$$v(y) = v(10000)w(p) \qquad (2)$$

where v is a value function, represented with $v(x) = x^{\alpha}$, and $w(p)$ is a subjective weight for probability p. In this study, $w(p)$ is represented by subjective value according to DbBS.

With the above assumptions, we estimated parameters for value function (i.e., α) and two parameters of the beta distribution whose CDF best explains the choice patterns in gambling task.

In the JE group, we estimated the best parameters based on the choice patterns for the 11 probabilities. In this estimation, we conducted a grid search; for α, from 0.04 to 1 with increments of 0.04 (i.e., 25 values); and for each of the two parameters of beta distributions, from 0.01 to 1 with increments of 0.01 (i.e., 100 values). Thus, in total, from 250,000 combinations of parameters, we searched the combinations of parameters, which explained the observed choice pattern best for every participant in the JE group.

For participants in the SE group, it was impossible to estimate their beliefs on uncertainty because they answered choices only for one gamble. Thus, we constructed a hypothetical participant who responded to 11 gambles (i.e., gambles for 11 probabilities), by the "SE" method with the following procedure. CEs for the 11 probabilities were constructed based on the data of the behavioral experiment. In particular, the CE at one probability was randomly sampled from normal distribution. Here, mean and standard deviation were determined by the data of the behavioral experiment (i.e., the data demonstrated in Figure 2). In these random samplings for 11 probabilities, we assumed that the hypothetical participant showed consistent choice patterns such that when $p_1 < p_2$, CE for p_1 (CE_1) and p_2 (CE_2) always satisfied $CE_{p1} \leqq CE_{p2}$. Thus, we estimated the response for 11 gambles using the SE method by the "identical person." With

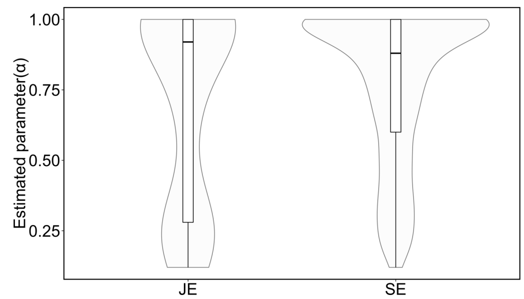

Figure 4. Distribution of estimated parameter for value function (α)

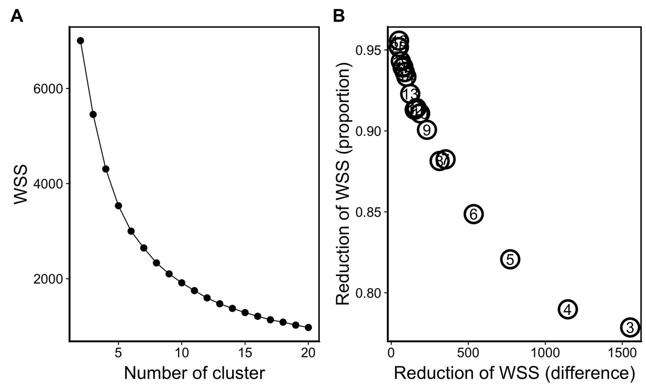

Figure 5. Results of clustering analysis. (a) Scree plot for within-cluster sum of squares (WSS) in K-means clustering. (b) Relationship between reduction of WSS (difference) and that in proportion. The number in the circle (e.g., n) indicates the reductions in WSS when the number of clusters increased from (n-1) to n.

these procedures, we constructed 1,000 hypothetical participants. For the data of the hypothetical participants, we estimate the best parameters for value function and beta distribution using a grid search as we did for the JE group.

In our parameter estimation, we evaluated the model fit using R^2. In the following analyses, we used the data wherein the model showed a good fit. Here, we set the criterion of "goodness" as $R^2 > 0.5$ (44 out of 47 data in the JE group and 787 out of 1000 data in the SE group satisfied this criterion).

Results of parameter estimations

Value function

Figure 4 shows the distribution of the estimated parameter of α for the JE and SE groups. As shown in the figure, the distributions were similar between the two group, and there was no significant difference ($w = 16328$, $p = .516$, Wilcoxon rank sum test). Thus, this result suggests that the different evaluations did not affect valuation of money.

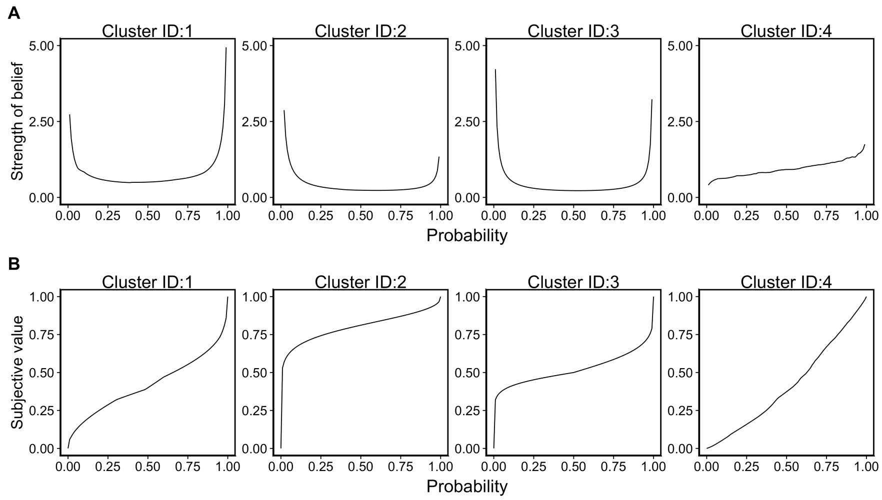

Figure 6. Median of strength of probabilistic belief (A) and subjective value (B) for the four clusters.

Table 2. Proportion of data categorized into the four clusters.

Group	Cluster 1	Cluster 2	Cluster 3	Cluster 4
JE	0.295	0.364	0.068	0.273
SE	0.287	0.159	0.475	0.079

Beliefs on uncertainty

Next, we examined participants' beliefs (i.e., estimated beta distribution) in detail with the following procedure. First, we clustered beliefs (i.e., shape of beta distribution) using probability densities. For the 831 data sets, the patterns of probability densities for 99 probabilities (1%, 2%, 3%,..., 97%, 98%, 99%) were clustered using the K-means method. We determined the number of clusters by considering the tradeoff between parsimony (i.e., as least clusters possible) and informativeness (i.e., as many clusters as required). Here, we calculated the within-cluster sum of squares (WSS) for each cluster and examined reductions in the WSS in terms of the difference and proportion of increasing numbers of clusters. Figure 5 shows the scree plot (a) and the relationship between the reduction in WSS in difference and proportion (b). We adopted four clusters based on their parsimony and informativeness.

We examined features of each cluster: median strengths of belief and median subjective values for 99 probabilities for each cluster. Figure 6 shows these results. The four clusters can be summarized as follows: For the clusters 1, 2, and 3, the probabilistic belief is "black and white" (i.e., deterministic). That suggests that a person refers to "winning" and "losing" gambles. The differences among the three clusters lie in whether a person is more optimistic (i.e., the belief in "winning" is stronger than that for "losing," Cluster 1), more pessimistic (i.e., the belief in "losing" is stronger than that for "winning," Cluster 2), or neutral (i.e., the belief in "losing" is as strong as that for "winning," Cluster 3). Cluster 4 has a different feature: the strength of belief is almost constant, suggesting that a person believes that the probability of winning gamble takes any probability (i.e., referring wide range of probability).

Then, we examined the proportions of data categorized into the four clusters for the two evaluation methods. Table 1 shows those results. Most data were categorized into Clusters 1, 2, or 3, which represented "black and white" belief. Those findings were generally consistent with the previous findings in Stewart et al. (2006) showing that people tend to often use extreme probabilistic expressions representing 0% and 100%. However, the most notable point was the proportion that was categorized into Cluster 4: more data from the JE group were categorized into Cluster 4 than from the SE group (p <.001, Fisher's exact test), suggesting that the participants (though "hypothetical participants") in SE referred to probabilistic information in a continuous way. These findings corroborated our prediction.

Discussion

In this study, we examined whether probability weighting shifts according to which evaluation method, JE or SE, was used in a gambling task. We found that the different evaluation methods induced different weighting. Furthermore, we analyzed our results using a cognitive model. The analyses indicated that differences in probability weighting for JE and SE were derived from a difference in probabilistic beliefs that people refer to in making decisions.

Previous studies have discussed changed preferences based on evaluation methods (JE, SE) but there has been little discussion about the process of making decisions under risk. One reason may be the difficulty of examining decision processes since researchers can obtain only one datum for each participant in an SE group, making model-based analysis highly difficult. In the present study, we proposed a new method to overcome such difficulties by constructing hypothetical participants using behavioral data. We believe that the proposed method makes a substantial contribution that helps clarify the difference in cognitive processes between JE and SE methods.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number 18H03501 for the second author and JP16H01725 for the last author.

References

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*(1), 129–166.

Honda, H., & Matsuka, T. (2014). On the role of rarity information in speakers' choice of frame. *Memory and Cognition, 42*(5), 768–779.

Honda, H., Matsuka, T., & Ueda, K. (2017). Decisions based on verbal probabilities: Decision bias or decision by belief sampling? In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 557–562). Austin, TX: Cognitive Science Society.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*(3), 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576–590.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62*(6), 1041–1062.

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

Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General, 132*(1), 23–46.

Stewart, N., Reimers, S., & Harris, A. J. L. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61*(3), 687–705.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.

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*(1), 7–11.