The impact of descriptions and incentives on the simultaneous underweighting and overestimation of rare events

Garston Liang (garston.liang@unsw.edu.au),

Emmanouil Konstantinidis (em.konstantinidis@gmail.com), Aba Szollosi (aba.szollosi@gmail.com) Christopher Donkin (christopher.donkin@gmail.com), & Ben R. Newell (ben.newell@unsw.edu.au) School of Psychology, UNSW, Sydney, NSW 2052, Australia

Abstract

We replicate and extend work demonstrating that choice and probability estimation can be dissociated through the coexistence of contradictory reactions to rare events. In the context of experience-based risky choice, we find the simultaneous *underweighting* of rare events in choice and their *overestimation* in probability judgement. This tendency persisted in the presence of accurate descriptions of rare event occurrence (Experiment 1), but was attenuated by incentivizing accurate probability estimates (Experiment 2). The implications of these results for popular models of risky choice are briefly discussed.

Keywords: Decisions from experience; probability estimation; risky choice; underweighting.

Introduction

Decision-making research often uses a description format to present risk information. In the experimental context of monetary gambles, participants are (usually) presented with two options. One option is deemed "safe" as it provides a sure payoff (e.g., a loss of \$3 with certainty), while the other option (e.g., a loss of \$4 with 0.8 probability; no loss otherwise) is deemed "risky" as it bears the risk of a rare event (i.e., the 0.2 probability of no loss). In the description format, the participant is explicitly given this risk information and makes a single choice between the two options. For the aforementioned gamble, Kahneman and Tversky (1979) found that participants preferred the risky option. This has been explained by assuming that participants choose as though they subjectively overweight the probability of the rare event. When the same gamble is presented in the gain domain, participants prefer the safe option (\$3 with certainty), consistent with the explanation of overweighting the 20% chance of receiving nothing.

In experience-based tasks, participants are presented with the same two choices without the aid of written descriptions. In order to learn about the outcome distributions, participants must repeatedly choose between the two options and observe the outcomes over successive rounds of choices. When risk information is presented in this format, choice preferences are consistent with *underweighting* the rare event. That is, participants prefer the safe option in the above example because the probability of the rare event (0.2 probability of no loss) is subjectively *underweighted*.

Probability judgement and underweighting

Fox and Hadar (2006) first proposed that erroneous probability judgements could be responsible for underweighting in experience-based choice. If participants judged the probability of the rare event to be lower than the objective probability, then underweighting in choice would reflect this erroneous judgement. To examine this judgement error hypothesis, a number of studies have asked participants to estimate the probability of the rare event at the end of the experiment (e.g., Hau, Plesak, Kiefer, & Hertwig, 2008). Using this retrospective method, studies have generally found probability estimates of the rare event to be accurate.

However, retrospective probes of this nature can create disparities between the information used during the task and judgements formed at the end of the task. Camilleri and Newell (2009) found that retrospectively generated probability estimates failed to predict participant choices during the experiment. One remedy for this is to prompt judgement probes throughout the task. In studies assessing awareness, Newell and Shanks (2014) suggested that assessments should be made as immediately as possible as to avoid forgetting and interference from other cognitive processes. This immediacy criterion can be applied to probability judgements as the large number of trials in experience-based tasks (e.g., 50 trials in Hau et al., 2008) may exacerbate any differences between retrospectively and immediately generated judgements.

The Coexistence Hypothesis

Barron and Yechiam (2009) used a novel trial-by-trial design to examine probability judgements in an experience-based task. In their Experiment 1, participants were repeatedly presented with a choice between a safe option (-3 points with certainty) and a risky option (-20 points with 15% probability; 0 points otherwise). Note that the expected value of both options was equal so a preference for the risky option would be indicative of underweighting the rare event. After an initial phase of choices alone, participants were prompted to estimate the probability of the rare event for the next trial following each choice.

Barron and Yechiam (2009) found that while participants preferred the risky option in their choices, consistent with underweighting the rare event, they overestimated the probability of the rare event (i.e., the probability of the -20 outcome). This inconsistency occurred at the individual level such that the majority of the participants (15/24) simultaneously demonstrated underweighting and overestimation of the rare event; a result that was named the coexistence hypothesis.

Barron and Yechiam (2009) also found evidence of opposing recency effects on choices and probability estimates. Following the observation of a rare event, participants were less likely to select the risky option. In these same trials though, probability estimates of the rare event were lower, demonstrating a reasoning process akin to the gambler's fallacy. This paradoxical result was called the contingent recency effect.

The coexistence hypothesis and the contingent recency effect are problematic for traditional decision models, such as the Two-Stage Choice Model (Fox & Tversky, 1998), because they present a dissociation between probability judgements and behavior. Most decision models like the Two-Stage model predict that choices can be mapped onto subjective representations of probability following a transformation according to the weighting function of Prospect Theory (Kahneman & Tversky, 1979). This makes the coexistence hypothesis an interesting anomaly given that overestimation is inconsistent with underweighting in choices.

Descriptions in decisions from experience

The existing literature separately compares decisions from experience to decisions from description. However, everyday decision making often utilizes a combination of both information sources (e.g., a doctor informed by both her clinical experience and empirical findings).

The limited number of studies examining decisions from experience in the presence of descriptions have produced inconsistent results. While some have found that the presence of descriptions influence choice in decisions from experience (Jessup, Bishara, & Busemeyer, 2008) others have contended that the descriptions are neglected (Lejarraga & Gonzalez, 2011). Aiming to resolve these contradictory accounts, Weiss-Cohen et al. (2016) found that participants predominantly relied upon experience to inform their choices but could be influenced by descriptions that provided novel information. This dovetails with the recent finding that the additional presence of descriptive summaries increased underweighting behavior in experience based tasks (Yechiam, Rakow, & Newell, 2015).

Taken together, these studies demonstrate that descriptions that explicitly provide probability information can influence choice in an experience-based task. However, less is known about how participants represent probability information when both sources of information are available. Given that risky choices are informed by probability judgements when each information source is presented separately, it is important to examine how individuals reconcile probability information in the presence of both descriptions and experience.

Experiment 1 examined the relationship between probability judgement and risky choice. We presented descriptions in an experience-based task that prompted trial by trial probability estimates similar to Barron and Yechiam (2009). We expected that participants given experience alone would show behavior consistent with the coexistence hypothesis (underweighting in choice and overestimation in judgement). Given that the descriptions explicitly stated the veridically experienced rate of rare events, we expected participants given description and experience to accurately estimate the rare event. However, as experience is primarily relied upon to inform choice patterns, we expected underweighting in choice would still emerge.

Experiment 1

Method

Participants Eighty undergraduate students ($M_{age} = 18.90$ years; SD = 1.66, 53 females, 1 other) from UNSW participated in exchange for course credit, and an incentive payment contingent on the participant's choices during the task (M = \$3.21 AUD, SD = 0.18).

Materials Participants were presented with two options associated with either a safe (S) or risky (R) distribution as follows:

Safe (*S*): -3 points with certainty

Risky (R): -20 points with probability 0.15; 0 points otherwise

The expected values were matched, and so a preference for the risky option would be indicative of underweighting. The risky distribution used random sequences of 120 outcomes, of which exactly 15% (18 outcomes) were rare events. Each sequence was presented to one participant in each condition.

Design Experiment 1 used a between-subjects design with 2 risk information conditions (n = 40). The description-experience (DE) condition completed the repeated choice task with descriptions that stated the outcome distribution of each option. For example, the description for the risky option read "15% chance of -20; 85% chance of 0". These descriptions remained visible for the duration of the experiment. The experience-only condition (E) completed the task without descriptions1.

Procedure Participants were randomly assigned to the E or DE condition and presented with instructions on a computer screen. Participants were told their show-up payment of \$5.00 AUD had been converted into 1000 points and that their task was to retain as many points as possible. They were prompted to make a choice between the S and R options on the screen (the locations of which were counterbalanced). Following each choice, full feedback for both the selected and forgone option was provided, which remained visible until participants proceeded to the next trial.

After 40 choice-only trials, participants completed an additional probability estimation task following each choice. Specifically, they were asked "What is the probability that -20 will appear in the next round?", and inputted an integer between 0-100 representing a percentage. Feedback for the current trial remained visible during the estimation task. This choice-then-estimation pattern was repeated for the remaining 80 trials after which the participants were debriefed and paid according to the rate of 2 points = \$0.01.

Results

Coexistence Hypothesis We found evidence of underweighting in choice in both the DE and E conditions.

¹A third condition with misleading descriptions called the conflicting descriptions experience condition has not been included here as it was not relevant to our current investigation.

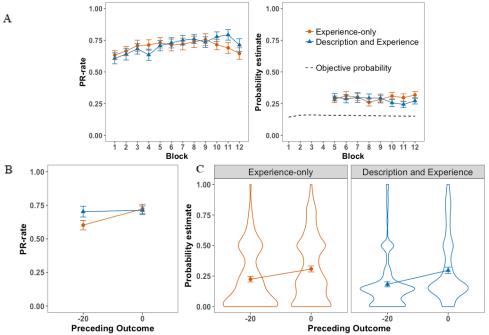


Figure 1. Risky choice and probability estimates data in Experiment 1. Error bars represent ± 1 SEM.
(A) Mean PR rates (left) and probability estimates (right) by block of 10 trials.
(B) PR rates as a function of preceding trial outcome.

(C) Violin plots of the mean probability estimates in each condition as a function of preceding trial outcome.

Overall, participants preferred the risky option as indicated by mean PR rates (the proportion of risky choices per block) being significantly greater than 0.5 in E condition, M = 0.70, t(39) = 6.50, p < .001, and the DE condition, M = 0.71, t(39) = 6.79, p < .001. As the expected values of the options were equal, this preference for the risky option (i.e., PR rates > 0.5) demonstrates that participants underweighted the rare event (Figure 1A, left panel).

We examined for differences between the conditions using a 2 x 12 mixed ANOVA with condition as the between factor and block (of 10 trials) as the within factor2. The main effect of condition was not significant, F(1,78) = .42, p = .84. This result indicates that we failed to find a significant difference in PR-rates between the experience-only condition and experience with explicitly stated descriptions.

Given that probability information was explicitly available, we expected that participants would accurately estimate the rare event. However, we found evidence of overestimation. Mean probability estimates were significantly greater than the objective probability of 0.15 in the E condition, M = 0.30, t(39) = 6.64, p < .001, and the DE condition, M = 0.28, t(39) = 5.65, p < .001. The effect of condition was non-significant, F(1,78) = .22, p = .64 (Figure 1A right panel).

Taken together, these results show the coexistence of inconsistent reactions to rare events in choice and probability judgement even in the presence of accurate descriptive information. Whilst underweighting suggests the rare event had less subjective impact than warranted on choices, overestimation suggests oversensitivity to the rare events' appearance. Moreover, the majority of participants (27/40 in

the E condition, 28/40 in the DE condition) exhibited this inconsistent pattern.

Contingent Recency Effect We separated participant responses into trials following a rare event (-20 outcome) and trials following a common event (0 outcome). This allowed us to assess for the impact of the most recently observed outcome, using a 2 x 2 mixed ANOVA with the outcome of the preceding trial as the within factor and condition as the between factor for both choices and probability estimates.

For choices, we found a significant interaction effect, F(1, 78) = 4.41, p = .04 (Figure 1B). This was qualified by a simple effects analysis which revealed a significant effect of preceding trial outcome for the E condition, F(1,39) = 8.25, p = .01, with a non-significant effect for the DE condition, F(1, 39) = .10, p = .76. This result shows that in the E condition people were less likely to select the risky option after observing a rare event compared to the common event (PR|-20 = .60 and PR|0 = .72), whereas in the DE condition this effect was not significant (PR|-20 = .70 and PR|0 = .71).

For probability estimates (PE), we found a significant effect of preceding trial outcome, F(1,78) = 22.98, p < .001 (Figure 1C). Averaged over conditions, participants estimated the probability of rare event to be lower following a rare event compared to following a common event (PE|-20 = .20 and PE|0 = .30). This is evidence of negative recency, which suggests that participants believed the chances of rare events in succession were unlikely. The effect of condition was not significant, F(1, 78) = 1.00, p = 0.32.

Taken together, these two results demonstrate an inconsistent reaction to the appearance of the rare event at least in the E condition. On the trials following a rare event, participants were less likely to select the risky option. However,

 $^{^{2}}$ In cases where the sphericity assumption has been violated, the Greenhouse-Geisser correction to *df* has been used.

participants simultaneously estimated the rare event to be less likely to occur. This contradictory pattern demonstrates a dissociation between choices and probability estimates, which replicates and extends the pattern observed by Barron and Yechiam (2009). (Note however, that Figure 1C shows different bimodalities in responding – a pattern that awaits further examination).

Experiment 2

The novel contribution of Experiment 1 is evidence of overestimation even in the presence of descriptions that explicitly state the probability of rare events. This suggests that overestimation emerges from factors related to experience.

Experiment 2 examined two hypotheses about the emergence of overestimation in experience-based choice. The first is that overestimation of the rare event is due to the anticipation of the loss of points. In Experiment 1, participants were incentivized on the outcomes of their choices and so each observation of the rare event was paired with a tangible loss of their incentive payment. This aversive experience may have led participants to overweigh rare events in memory. If this is the case then accuracy might improve if probability estimation was decoupled from experiencing the consequential outcome.

The second hypothesis is that overestimation arises from psychophysical limitations with probability processing. The suggestion here is that although individuals may be proficient in tracking the frequency of events (e.g., Hasher & Zacks, 1979), they may have difficulties in expressing this information as probabilities (cf. Gigerenzer & Hoffrage, 1995). This explanation would predict that overestimation is not affected by consequential outcomes but arises due to an inherent incapability of accurately estimating probabilities. Therefore, even if risky choices were divorced from any loss of points, overestimation would still occur.

We tested these hypotheses by separating consequential choices from probability estimation using different incentive schemes. Specifically, participants were incentivized on i) choices only, ii) only the accuracy of probability estimates, or iii) both choices and the accuracy of probability estimates.

Method

Participants In Experiment 2, 132 students (M_{age} = 19.52 years; *SD* = 1.64, 100 females) participated in exchange for course credit, and a performance-based incentive payment.

Materials As in Experiment 1.

Design Experiment 2 used a between subjects design with 3 groups (n = 44) differing in incentive structure.

The choice-incentive group was incentivized on the outcomes of their choices, replicating the E condition from Experiment 1. The probability-incentive condition was incentivized on the accuracy of their probability estimates. In this condition, participants did not choose between the options but instead pressed a separate button that revealed the outcomes from both options simultaneously. The outcomes were financially inconsequential and their task was to track the

outcomes in order to accurately estimate the rare event. Accuracy was calculated as the percentage point deviation from the experienced probability of rare events on each trial. These deviations were tallied at the end of the experiment rather than during the experiment. This was done to avoid giving any feedback about the accuracy of the estimates which could have influenced responses during the probability estimation task. In the dual-incentive condition, participants were incentivized on both the outcomes of their choices and the accuracy of their probability estimates.

Procedure Participants were provided with instructions on a screen that explained the incentive scheme and the objectives of their respective conditions. Across conditions, participants were given a \$5.00 show-up payment from which an amount would be deducted contingent on their performance in the task.

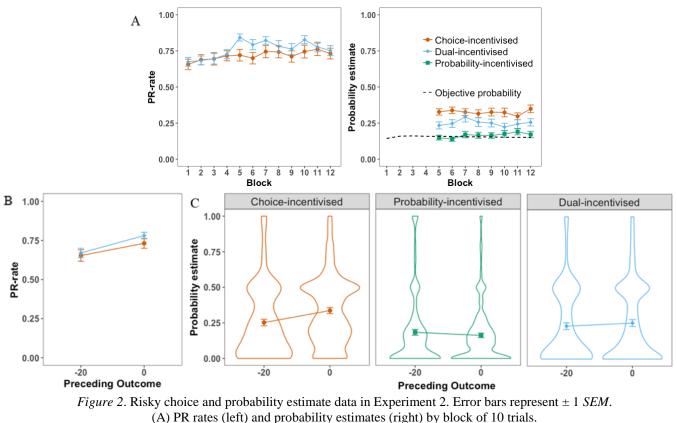
In the initial stage, participants in the choice-incentive and dual-incentive conditions made repeated choices while those in the probability-incentive condition tracked the outcomes of both options. After 40 trials, all participants also completed the probability estimation task on each trial. Participants were then debriefed and paid accordingly. Choice incentive participants converted their remaining points at a rate of 2 points = 0.01(M = \$3.19, SD = .19). Probability-incentive participants were penalized 0.05 cents per percentage point deviation (M =4.44, SD = .31). For the dual-incentive participants, their show-up payment was divided in half with each half paid according to different incentive structures. The "choice" half of the payment was calculated by converting their remaining points at a rate of 1 point = 0.5 cents. The "probability estimate" half was penalized at 0.025 cents per percentage point deviation.

Results

Coexistence hypothesis PR-rates were again significantly greater than .5 in the choice-incentive condition, M = .72, t(43) = 7.28, p < 0.001, and the dual incentive condition, M = .76, t(43) = 12.98, p < 0.001 (Figure 2A, left panel). This suggests participants underweighted the rare event in their choices. A mixed 2 x 12 ANOVA, with condition as the between and block as the within factor, found that the effect of condition was not significant, F(1, 86) = 1.46, p = .23.

Participants overestimated the rare event in the choiceincentive, M = 0.33, t(43) = 8.99, p < .001, and dual-incentive conditions, M = 0.26, t(43) = 4.50, p < .001, shown by mean probability estimates significantly greater than the objective probability of 0.15 (Figure 2A, right panel). By comparison, mean estimates in the probability-incentive condition did not differ significantly from the objective probability of 0.15, M =0.16, t(43) = 1.09, p = .28). Using a 3 x 12 mixed ANOVA, we found a significant main effect of condition, F(2, 129) = 17.51, p < 0.001. Post-hoc Scheffé comparisons showed that all groups significantly differed from each other. The interaction effect was not significant, F(10.85, 699.63) = 1.14, p = .33.

In summary, mean probability estimates were highest in the choice-incentive condition where participants were incentivized for the outcomes of their choices alone. However,



(B) PR rates as a function of preceding trial outcome.

(C) Violin plots of the mean probability estimates in each condition as a function of preceding trial outcome.

if participants were additionally incentivized for accuracy as in the dual-incentive condition, estimate accuracy was improved. Mean probability estimates were most accurate when consequential outcomes were removed all together.

Taken together, these results replicate the coexistence hypothesis of underweighting in choice and overestimation in probability judgement. We found that 34/44 participants in the choice-incentive condition, and 26/44 participants in the dual-incentive condition demonstrated this pattern of responding.

Contingent Recency Effects A 2 x 2 mixed ANOVA found a significant effect of preceding trial outcome for risky choices (Figure 2B). This showed that averaged over conditions, PR-rates were lower following a rare outcome than following a common outcome (PR|-20 = .66 and PR|0 = .76), F(1, 86) = 18.91, p < .001.

The mean probability estimates contingent on the preceding trial outcome are presented in Figure 2C. We used a 3 x 2 mixed ANOVA with preceding trial outcome as the within factor to examine for recency effects. A significant effect of preceding trial outcome was found, F(1, 129) = 4.42, p = .04, which was qualified by a significant interaction between condition and preceding trial outcome, F(2, 129) = 5.34, p = .01. Simple effects analysis revealed probability estimates in the choice-incentive condition were lower after observing a rare outcome than after observing a common event (PE|-20 = .25 and PE|0 = .34), F(1, 129) = 13.33, p < 0.001.

By comparison, the effect of preceding trial outcome was not observed in the probability-incentive condition (PE|-20 = .18 and PE|0 = .16), *F* (1,129) = 0.90, *p* = 0.35, or the dual-incentive condition (PE|-20 = .23 and PE|0 = 0.25), *F* (1,129) =

0.88, p = 0.35. Taken together, the negative recency effect of preceding trial outcome was only found when participants were not incentivized to estimate accurately. (Note again however, that Figure 2C shows bimodality in responding – clustered around 50 and 0 – a pattern that awaits further examination).

General Discussion

Across two experiments, we found the coexistence of underweighting in choice and overestimation in probability judgements at the individual level. Furthermore, inconsistency is evident in the trials immediately following rare events. Experiment 1 replicated the coexistence hypothesis in the presence of accurate descriptions. We failed to find a difference in choices and probability estimates between participants that received descriptions and those that did not. Experiment 2 used incentive schemes to show that overestimation emerges in the presence of consequential outcomes. We postulate that attention to the probabilities attenuated the degree of overestimation.

The results of our experiment suggest consequential outcomes biased attention away from probability tracking. Kahneman (1973) defined attention as a limited resource that is allocated according to the demands of the task. Overestimation may have occurred in anticipation of the loss of points associated with the rare event, driving attention towards the outcomes themselves. With respect to the dual-incentive condition, the presence of consequential outcomes in the choice task meant fewer attentional resources could be allocated to probability tracking. This competition for attentional resources between the two tasks would explain why the degree of overestimation was attenuated, but still remained in the presence of consequential outcomes.

Our overestimation results are incompatible with the judgement error hypothesis (Fox & Hadar, 2006). Therefore, the coexistence of overestimation and underweighting suggests that probability judgement and choice may reflect separate cognitive processes. The distinction between choice and judgement resembles the comparison between *online* and *memory-based* strategies. Hastie and Park (1986) distinguished between two types of strategies based on how information is processed to form a judgement. Online strategies involve step-by-step information processing whereby a judgement is continually updated with new information. By comparison a memory-based strategy involves a discrete instance in which all relevant information is recalled from memory to form a judgement (Haberstroh & Betsch, 2002).

We propose that the choice process resembles an online strategy that involves continually updating a small mental sample of outcomes with recently observed outcomes. This is compatible with the explanation that underweighting emerges from small mental samples of outcomes (i.e., calculating the expected value of the last 5 observed outcomes, cf. Erev & Roth, 2014). Concurrently, probability judgements resemble a memory-based strategy where aversive rare events are overweighted in memory, resulting in overestimation (e.g., Tversky & Kahneman, 1973).

Separate processes for choice and judgement would be consistent with our findings of the contingent recency effects. In trials following the rare event, participants were less likely to choose the risky outcome yet, they paradoxically estimated a lower probability of the occurrence of the rare event. An online strategy for choices would involve a trial-by-trial updating process that incorporates each newly observed outcome into the decision process. Whilst the risky option usually provided the more attractive outcome (i.e., no loss of points), the occasional appearance of the rare event meant that the small sample from the risky option was momentarily less attractive than the safe option. This explains the reduced tendency to select the risky option after observing a rare event. In these very trials, a memory-based strategy that employs the gambler's fallacy would explain the lower probability estimates. Given a rare event on the preceding trial, participants may have reasoned that "lightning does not strike twice". Therefore, the gambler's fallacy may have served as a memory heuristic to simplify the more cognitively demanding memory-based estimation process.

In summary, we have shown the impact of descriptions and incentives on the simultaneous overestimation and underweighting of rare events.

Acknowledgments

This work was supported by the Australian Research Council (DP140101145; DP160101186).

References

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the

contingent recency effect. Judgment and Decision Making, 4, 447-460.

- Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making*, *4*, 518-529.
- Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, 111,10818-10825.
- Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*, 159-161.
- Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, 44, 879-895.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704.
- Haberstroh, S., & Betsch, T. (2002). Online strategies versus memory-based strategies in frequency estimation. *ETC. Frequency Processing and Cognition*, 205-220.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356-388.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 258-268.
- Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21, 493-518.
- Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, 19, 1015-1022.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 263-291.
- Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. Organizational Behavior and Human Decision Processes, 116, 286-295
- Newell, B.R. & Shanks, D.R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, 37, 1-63
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.
- Weiss-Cohen, L., Konstantinidis, E., Speekenbrink, M., & Harvey, N. (2016). Incorporating conflicting descriptions into decisions from experience. *Organizational Behavior and Human Decision Processes*, 135, 55-69.
- Yechiam, E., Rakow, T., & Newell, B. R. (2015). Super-underweighting of rare events with repeated descriptive summaries. *Journal of Behavioral Decision Making*, 28, 67-75.