Risky Intertemporal Choice with Multiple Outcomes and Individual Differences

Daniel G. Wall (dwall@andrew.cmu.edu)[1],
Gretchen Chapman (gchapman@andrew.cmu.edu)[1],
Pernille Hemmer (pernille.hemmer@psych.rutgers.edu)[2]
[1] Social and Decision Sciences, 5000 Forbes Ave, Pittsburgh, PA 15213 USA
[2] Psychology, 152 Frelinghuysen Road, Piscataway, NJ 08854 USA

Abstract

Risk and delay co-occur. Intertemporal choices are rarely certain; risky choices are rarely atemporal. Behavioral evidence suggests that risk and time are entangled: time discounting is different for risky outcomes than for riskless outcomes. A prominent model of risky intertemporal choice (Baucells & Heukamp, 2012) combines risk and delay into psychological distance. It predicts that risk and time will be entangled for outcome risk (risk with one zero outcome and at least one positive outcome) but not for amount risk (risk with three or more positive outcomes) unless assuming non-cumulative probability weights. We show that BH does not quantitatively fit risky intertemporal choices better than a model assuming risk and time are independent. Many participants were best fit by a random response model. The functional form for risky intertemporal choices is difficult to detect. While risk and time are entangled, they do not seem to be evaluated as psychological distance.

Keywords: Risky Intertemporal Choice; Choice Modeling; Individual Differences, Latent Mixture Model

Introduction

Decisions often involve a trade-off realized over different time periods. When choosing an age to retire you can wait to receive Social Security payments of $1200 a month or claim early and receive only $1000 a month (Knoll, Appelt, Johnson, & Westfall, 2015; Schreiber & Weber, 2016). Many real-world choices, beyond the delay to receipt, have an element of risk in the outcome. For instance, when investing in an individual retirement account (IRA), you know that the total value at the time of your retirement is uncertain. However, most models of risk and time treat each aspect independently. This might be why there is mixed evidence for the predictive power of intertemporal choices on real-world decision making (Arfer & Luhmann, 2017). To better predict real world decision making, models should acknowledge the co-occurrence of risk and delay.

Recent work has proposed an integrated model of risk and time (Baucells & Heukamp, 2012), but this model implicitly ignores the fact that risk is rarely binary—you are unlikely to encounter a prospect which has a 50% chance of $100 and a 50% chance of $0. Your IRA can take on a lot of different values in the future. Further, as is often the case, a model is tested via the qualitative behavior effect it produces, not how well it recovers parameters and predicts actual choices.

The treatment of risk and time as separate is further perpetuated in experimental work. As opposed to measuring risk and time together, the dominant paradigm for investigating temporal preferences is choices between two certain (i.e., without risk) monetary amounts, each received at a different point in time. For example, participants might be asked to make a choice between receiving $100 today or $110 in 4 weeks. Adding risk can be achieved by specifying a 50% probability for all monetary amounts (Ahlbrecht and Weber 1997; Andreoni and Sprenger 2012; Baucells and Heukamp 2012). This increased realism, however, it also alters discounting.

Consider the immediacy effect, a phenomenon in which people display a strong preference for immediate outcomes. People prefer $100 today to $110 in 4 weeks, but if you push both options back 26 weeks, people prefer $110 in 30 weeks to $100 in 26 weeks. The power of immediacy, however, diminishes in the presence of risk. For instance, Weber and Chapman (2005) demonstrated that adding outcome risk—e.g. specifying a 50% probability for all monetary amounts—attenuates the immediacy effect. In other words, the power of now may in part be due to the certainty of now. Related work has reached a similar conclusion: people discount the future differently if it is risky as compared to if it is certain (Andreoni & Sprenger, 2012; Hardisty & Pfeffer, 2015). Taken together, these results suggest that risk and time are entangled.

In this paper we measure the entanglement of risk and time by fitting multiple models of risk and time to choices, including a random response model. We further allow for individual differences in risk time entanglement. Entanglement has only been observed in aggregate; however, this may be due to a subpopulation which has entangled preferences. Lastly, to increase realism we include choices which have multiple positive outcomes.

In the next section we discuss two broad categories of risky intertemporal choice models: one category assumes risk and time are disentangled and the other assumes they are entangled. For the entangled model, we focus on the Baucells and Heukamp (BH) model which reduces to a disentangled model. In order to see how these models deal with risk with multiple outcomes, we extend both the entangled and disentangled models with cumulative and non-cumulative prospect theory. We then fit the 4 models and a random response model to choices, allowing each participant to be fit by a unique model. Our results suggest that risk and time – while entangled – are not unified under psychological distance—an assumption of the BH model.

Two Classes of Risky Intertemporal Choice Models

Models of risky intertemporal choices fall into two broad categories. The first category assumes that risk and time are evaluated independently and therefore involves a two stage process: 1) calculate the certainty equivalent of the risky option, then, 2) discount it (Abdellaoui, Diecidue, & Öncüler, 2011; Öncüler, 2000). Evaluating time and risk separately assumes that they are independent of one another, meaning that, contrary to the evidence presented above, risk should not affect temporal discounting. The second category, which evaluates time and risk together, in a single step, assumes that they are entangled and therefore are dependent on one another (Baucells & Heukamp, 2012). This combined evaluation assumes that time and risk are evaluated in concert, meaning that the effect of risk on value is dependent upon the delay to receipt and vice versa.

For the entangled category of models, we use the BH model because it has only three free parameters and has a psychological interpretation for risk-time entanglement. Specifically, it assumes that both risk and time increase psychological distance—which is subadditive. Further, while the model was originally presented as a quantitative model of risky intertemporal choice, it was not fit to choices. More practically for our purposes, BH reduces to a common probability weighting model Prelec (1998), for immediate prospects, and a common time discounting model, Ebert and Prelec (2007), for riskless prospects.

For the disentangled category of models, we use a two-stage model that combines the Prelec probability model with the Ebert and Prelec time discounting model and evaluates risk and time independently, i.e. the effect of delay is independent of risk and the effect of risk is independent of delay. While the BH model also combines the Prelec model with the Ebert and Prelec model, it evaluates delay and risk together, e.g. the effect of delay on value depends on how much risk and the effect of risk on value depends on delay. More simply, the BH model assumes risk and time are entangled, while the Prelec then Ebert and Prelec model assumes risk and time are disentangled.

Disentangled Model Stage 1: Prelec Probability Model of Probability Weighting

The first stage of our two stage model calculates, given an objective probability, the decision weight of an outcome. The decision weight of an unlikely event ($p = .01$) is generally overweighted, while the perceived probability of a likely event ($p = .99$) is generally underweighted. Prelec (1998) outlines a probability weighting model which defines the weighted, or perceived, probability *w(p)* of an objective probability p as

$$w(p) = \exp((-(-\ln p)^{\delta_P})) \qquad (1)$$

Where p is the probability of the option and the δ_P parameter is the distortion of probabilities. For $\delta_P = 1$ there is no distortion of probability, but for $\delta_P = 0.1$ probabilities are very distorted, with extreme overweighting of low probabilities, $w(.01) = 0.31$, extreme underweighting of high probabilities, $w(.99) = 0.53$. The dollar amount x is then multiplied by the weight, $w(p)$, to get the value, $\mathrm{V}(x, \delta)$, of the gamble. For ease of exposition, we are assuming linear utility[1]. The Prelec probability weighting model calculates, ignoring delay, the risk adjusted value of a prospect. The second stage of our two stage model—the Ebert and Prelec model of time discounting—then discounts that value .

Disentangled model stage 2: The Ebert and Prelec Model of Time Discounting

The Ebert and Prelec (2007) model of time discounting has a similar functional form as the Prelec model, but instead of accounting for deviations from linear probability weighting, it accounts for deviations from exponential discounting. Specifically, the Ebert and Prelec model calculates the discount factor $d(t)$—the factor by which a present value is reduced moving t days into the future—by:

$$d(t) = \exp(-(r_{daily}\, t)^{\delta_{EP}}) \qquad (2)$$

Where t is the delay, r_{daily} is the daily discount rate, and δ_{EP} is the index of time sensitivity. When $\delta_{EP} = 1$, the Ebert and Prelec model reduces to exponential discounting. When $\delta_{EP} < 1$ the model predicts increased discounting for delays in the near future and decreased discounting for delays in the far future—more simply, it can account for the immediacy effect.

Disentangled Model: Combined Prelec and Ebert-Prelec (PEP) model

The discount factor from the Ebert and Prelec model—Disentangled model stage 2—is multiplied by the risk adjusted value—Disentangled model stage 1—to calculate the utility of the delayed prospect. Specifically, the probability adjusted discount factor, α, of a delayed prospect is the product of the Prelec probability weight and the Ebert and Prelec discount factor: $\alpha = w(p)d(t)$. The value of a delayed prospect is: αx. This combination of the Prelec probability weighting model and the Ebert-Prelec discounting model is a two stage model of delayed prospects which we call the PEP model. The PEP model assumes *disentangled* risk and time; it is a foil to the predictions of the BH Model.

Entangled Model: Baucells and Heukamp

In contrast to the two stage PEP model, Baucells and Heukamp (2012) assumes that risk and time are entangled. BH combines the Prelec and Ebert and Prelec models into a one stage model where risk and time form a single distance function, which captures the psychological distance to the outcome. Increasing the amount of time to an option increases psychological distance; similarly, decreasing the probability of receiving an outcome increases psychological distance (Trope & Liberman, 2010). This psychological distance function is subadditive, such that the combination of risk and time produces an effect on psychological distance

[1] The utility function is similar for both time and risk (Luckman et al., 2017)

that is less than the linear combination of risk and time by themselves.

For a delayed prospect, the BH distance model calculates probability adjusted discount factor, α, as:

$$\alpha = \exp\left(-\left(-\ln p + r_{daily}\, t\right)^{\delta_{BH}}\right) \quad (3)$$

where p is the probability with which it will be received and t is the delay until receipt. The parameters for calculating utility are r_{daily}—the probability discount rate which measures the trade-off between probability and time delay—and δ_{BH}—the sensitivity to psychological distance. Higher δ_{BH} values indicate less sensitivity to distance. For lower δ_{BH} values distance is more subadditive. The BH model reduces to the Ebert and Prelec time discounting model for certain, i.e. $p = 1$, intertemporal choices and reduces to the Prelec Probability Weighting function for atemporal, i.e. $t = 0$, risky choices (Baucells & Heukamp, 2012; Toubia, Johnson, Evgeniou, & Delquié, 2012).

In order to keep terms consistent across models, we refer to the δ parameter as the *deviation parameter.* In the Prelec model it refers to the deviation from linear probability weighting, in the Ebert and Prelec model it refers to the deviation from exponential discounting, and in the BH model it refers to the diminishing sensitivity between risk and time, e.g., how much does risk and time deviate from risk alone and time alone.

Two Types of Risk, Two types of decision weights

In pilot studies, we found that BH does not predict that amount risk alters discounting unless we assume noncumulative decision weights. This, however, seems implausible because, for amount risk the total probability weight can be greater than 1, leading to violations of dominance. In order to test if noncumulative probability weights actually fit choices better, here we fit all four models—BH Cumulative, BH NonCumulative, PEP Cumulative, PEP NonCumulative—as well as a random response model to choice data.

A key element of Prospect theory (Kahneman & Tversky, 1979) is the overweighting of small probabilities and the underweighting of large probabilities. For prospects with two positive values—½ chance $25 and ½ chance of $100 — probability weighting can lead to violations of stochastic dominance. To correct for this this, Kahneman and Tversky propose editing rules—the prospect assuredly pays out $25 and has a ½ chance of receiving an additional amount. However, this editing rule does not easily apply to prospects with, say, four (which is what we use in our experiment) positive outcomes (Gonzalez & Wu, 1999). Therefore, we estimate *noncumulative* decision weights without an editing rule. For the following prospect, a 1/6 Chance of $125; a 1/6 Chance of $120; a 1/6 Chance of $115; 1/2 chance of $0, we simply calculate a decision weight for 1/6th and multiply that by the dollar amount and take the sum.

While *noncumulative* decision weights can lead to violations of stochastic dominance and seem to be prima facie incorrect, there is reason to believe that delay alters probability weighting. Specifically, the form of decision weighting may be entangled with time. If people perceive delay as adding risk, they should weight probabilities differently when delay is present. As previously mentioned, when combined with the BH model, noncumulative decision weights predict behavioral patterns in risky intertemporal choices, but cumulative decision weights do not.

In order to deal with violations of stochastic dominance without ad hoc editing rules, Tversky and Kahneman (1992) created *Cumulative* prospect theory. What follows is Gonzalez and Wu's (1999) intuitive specification of cumulative decision weights. First, the options are placed in order according to decreasing absolute value. Formally, where X_1 is the dollar amount for the first option and p_1 is the probability of receiving the first option, a prospect is represented by $(X_1, p_1; \ldots; X_n, p_n)$ where $|X_i| > |X_{i+1}|$ and all X's are on the same side of the reference point. The value of a prospect is the following:

$$w(p_1)u(X_1) + \sum_{i=2}^{n}\left[w\left(\sum_{j=1}^{i} p_j\right) - w\left(\sum_{j=1}^{i-1} p_j\right)\right]u(X_i) \quad (4)$$

where u is the utility function, which in our case is linear.

Taking the atemporal prospect outlined in the noncumulative section, a cumulative probability weighting proceeds as follows—remember we assume linear utility. Since the items are already in decreasing order according to absolute value, "1/6 chance of $125" would be have a decision weight of $w(1/6) * \$125$. The weight of $120 would be $[w(1/3) - w(1/6)] * \$120$, which is the decision weight of the probability of receiving at least $120 minus the decision weight of the probability of receiving more than $120. And similarly to $125, the weight of $115 would be $[w(1/2) - w(1/3)] * \$115$, e.g. the probability weight of receiving at least $115 minus the probability weight of receiving more than $115. This sets an upper bound on the sum of decision weights at 1. In the two stage model, the weighting function, w, is the Prelec probability weighting function, but for the one stage model it is the BH model.

For convenience we call prospects with multiple positive dollar values amount risk. For instance: a 1/3 chance of $100; a 1/3 chance of $120; a 1/3 chance of $1000 has amount risk (Blackburn & El-Deredy, 2013; Hardisty & Pfeffer, 2015). Moreover, the real world presents prospects with many eventualities, our models should account for this. Outcome risk, however, is when there are only two outcomes and one is a zero. For instance, a ½ chance of $100 and a ½ chance of $0 is outcome risk. Outcome risk (2 outcomes, one is zero) is a very special case of the broad class of lotteries. This is admittedly a fuzzy definition, but it's a useful demarcation which has been used by other researchers studying risky intertemporal choices (Blackburn & El-Deredy, 2013; Hardisty & Pfeffer, 2015). A two outcome prospect is as simple as it gets and when one of the two options is 0 there is no need for editing rules.

In addition to investigating amount and outcome risk individually, we also combined amount and outcome risk. For instance: a 1/4 chance of $100; a 1/4 chance of $120; a

1/4 chance of $1000; a 1/4 chance of $0 has amount and outcome risk.

Method

Participants

A sample of 18 MTurk participants made the 120 choices outlined above. Two participants missed more than 2 of the check questions and were removed from the data[2]. In order to incentivize honest answers, participants were told that one of their choices would be paid out for real, except that the value would be 1/100 of the nominal value in the survey. To ensure attention, participants were told that their bonus payment was contingent on answering the check choices correctly.

Stimuli and Procedure

To create stimuli, we followed the methodology of Erev et al., (2014) which sampled from distributions for probabilities and dollar amount. First to eliminate participant's editing out risk common to both options, the smaller present option was always certain. Second, the algorithm created stimuli with expected values of about $100. Third, and most importantly the stimuli had varying delays. The 120 stimuli created from the algorithm are at https://osf.io/sv5xn/; and the R code used to create the stimuli is at https://osf.io/xdpmz/. An example choice with delay, amount, and outcome risk is: 1) $87 for certain today 2) A 14% chance of $235 in 60 days; A 6% chance of $775 in 60 days; A 18% chance of $87 in 60 days; A 62% chance of $0 in 60 days. This design had at minimum 10 check choices for the atemporal arisky choices. Since the algorithm had random values for some choices, there was an uncertain number of additional check questions. The random number seed yielded an additional 7 check choices for arisky intertemporal choices which had a dominated option, for a total of 17 check choices—choices in which one item asymmetrically dominated the other option. In order to have a completely within subjects design, all 120 stimuli were presented in a random order to each participant.

To validate our stimuli, we performed two forms of parameter recovery. First we followed the methodology of Broomell and Bhatia (2014) which demonstrates the stimuli's ability to recover parameters from a large parameter space. These stimuli were comparable to the other stimuli sets in Broomell and Bhatia.

The second parameter recovery followed Nilsson, Rieskamp, and Wagenmakers (2011). Specifically, we generated synthetic data for 120 participants and used these data in the hierarchical Bayesian mixture model outlined in the next section. The model recovered both the generative parameters and the risky intertemporal choice model which generated the data. Taken together, the stimuli created do a good job of recovering parameters and distinguishing between risky intertemporal choice parameters.

Results

Mixture Model

When disaggregated, it is possible that for some people risk and time are entangled, but for other people they are disentangled. In order to see which of the five models—BH Cumulative, BH NonCumulative, PEP Cumulative, PEP Noncumulative, and random responding—fit individual participants best, we used a latent mixture model. Latent mixture models—aka latent class models—can be used to determine which model best fits each individual participant's data.

We used a hierarchical modeling framework. Each participant's mean δ was drawn from a population with a mean and standard deviation, and then was given 0-1 support via the probit transformation. The same procedure was used for each participant's r parmeter. For the logit scale sensitivity parameter, as opposed to the probit transformation, we used a lognormal transformation.

Of note is the mixing parameter, which is a categorical variable that indexes which model fit a person's data the best. The mixing parameter was not hierarchical. The mixing parameter has the following structure: 1st group is the cumulative BH, 2nd group is noncumulative BH, 3rd group is cumulative pep, 4th group is noncumulative pep, 5th group was a random response model, which ignored all other information and essentially flipped a coin for each choice. The fifth, random response, model was included to ensure that a participant who was not fit by any of the four models did not contaminate the hierarchical parameter values. More importantly, the random response group indicates the number of participants who have a generative process which is not well estimated by the other models.

Hierarchical Bayesian Latent Mixture Modeling

Figure 1 shows that most participants were fit nearly equally well by all models. Further, five of the participants were better fit by a model of random responding than any of the four models. While it would be interesting to see if those who were fit by a random response model showed entanglement, the stimuli set in our study is not conducive to a test of behavioral phenomenon.

[2] While this is an admittedly small sample, there are a larger number of individual level choices. Rieskamp 2008 used only 30 participants to fit a Cumulative Prospect Theory model. Further, we ran another sample with slightly modified stimuli, and the results were nearly identical.

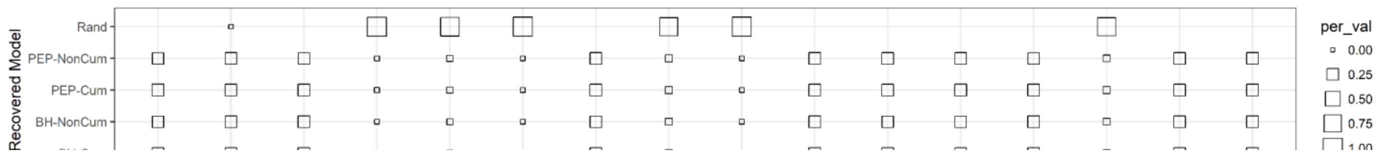

Figure 1. . Probability of each participant being fit best by the four models and random responding. The size of the box indicates the probability that model fit the participant best.

As seen in Figure 2, the hierarchical mean of delta was almost always near 1. When the delta parameter is higher, the four models collapse and become nearly identical. The BH model becomes: $\exp(-(-\ln p_i + r_n t_i))$ and the PEP model becomes $\exp\big(-(-ln p_i)\big) \times \exp(-r_n t_i)$, and via the product rule, these two models are the same. Since there is no probability distortion, $\delta = 1$, cumulative decision weights reduce to noncumulative decision weights. It is tempting to rerun the analyses with an unbounded delta parameter, however, this makes the assumption that people could *underweight* low probabilities and *overweight* high probabilities. While the description-experience gap predicts underweighting of small probabilities, there is reason to believe this prediction is an artifact due to a large number of possible decision weights that fit experienced prospects (Broomell & Bhatia, 2014). A model that predicts anything predicts nothing.

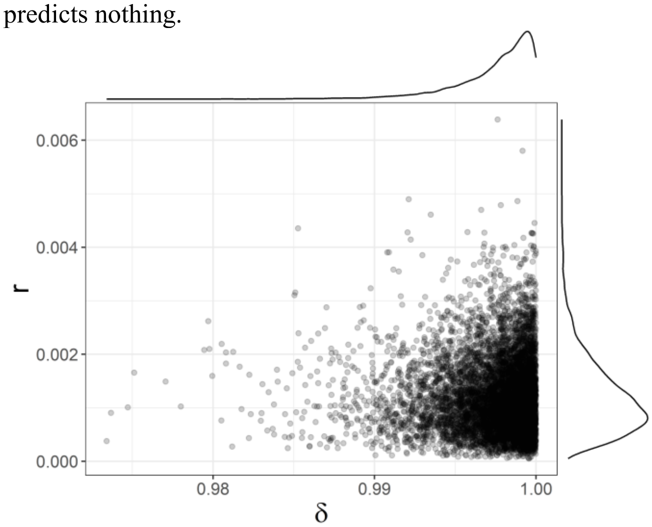

Figure 2. Joint posterior of the hierarchical means of δ and r. Each point is a sample draw from the Markov Chain.

Un-Mixing the mixture model

In order to determine if the high δ in the latent-mixture model was due to certain choices, we fit models to various subsets of the data. Specifically, for the choices which were simply intertemporal choices, we fit the Ebert and Prelec model—all four models reduce to the Ebert and Prelec model for arisky temporal choices. For risky items which are atemporal, we fit the cumulative Prelec model—the cumulative BH model reduces to the Prelec model for atemporal risky choices. For all of the remaining choices, e.g. temporal risky choices, we fit the cumulative BH model.

For arisky temporal choices, the hierarchical Bayesian model converged—all Rhats < 1.1 and visual inspection of the traceplots confirmed that the posterior was in a stationary distribution. Further the mean value for the δ parameter was 0.86 with a 95% credible interval of [.55 - .99]. This estimate was higher than previous estimates of the δ_{EP} parameter .18-.6 (Ebert & Prelec 2008). This may be due to participants adopting a different heuristic when answering risky intertemporal choices than when they answer solely intertemporal choices. It is worth considering that participants only saw 13 pure intertemporal choices, meaning that the data do not have a chance to overwhelm the diffuse prior on δ_{EP}. With a narrower prior, the posterior δ_{EP} would be closer to 1. The δ_{EP} parameter is high in our data for arisky intertemporal choices, meaning that people's discounting is fit by a discount function which is close to exponential discounting.

For atemporal risky choices, the hierarchical Bayesian model for the cumulative Prelec model converged—Rhats < 1.1 and traceplots indicated convergence. The mean value of the delta parameter was .99 with a 95% credible interval of [.98, .999]. This value means that people's distortion of probability is very close to linear. However, we must exercise caution in interpreting the parameter values because the model posterior predicted participant's actual choices less than 50% of the time. As a secondary check we also fit a model with just atemporal outcome risk choices and the model posterior predicted choices 37% of the time. We refit the exact same model on the simulated parameter recovery data with $\delta_P = .6$ and the model recovered the parameters well and had a 92% hit rate.

For the remaining, risky intertemporal choices, the model converged as well. The mean δ value was .985 with a 95% credible interval of [.90 - .9999], this suggests that the delta value is quite high for these data. Taken together, it seems as if the empirical delta is too high to fit the latent mixture model and distinguish between different generative processes.

As another check, we redid the Nilsson et al. (2011) parameter recovery outlined above, except with a generating delta of .95. The results were nearly the same as the empirical data, each participant was fit equally well by all models, and no participants had more than 50% of the mixing probability in their true generative process.

Discussion

In this paper we aimed to test the entanglement of multiple forms of risk and delay. Specifically, we were interested in individual differences in entanglement and how risk with multiple positive outcomes related to entanglement. As opposed to looking at behavioral phenomenon, we fit these models to actual choices. Specifically, we fit a mixture model which allowed participants to be fit best by either the BH Cumulative, BH NonCumulative, PEP Cumulative, PEP NonCumulative, or a random response model. Two assumed entanglement of risk and time and two assumed disentanglement of risk and time. We were unable to determine which of the four discounting models fit participant's data the best. Nearly 1/3 of participants were fit better by random responding than by any other model.

Taken together, this suggests that while the behavioral effects are consistent with a BH model using noncumulative decision weights, the model is actually not a good fit to the data. This work implores caution: using solely behavioral effects as a test of models can be insufficient. To determine the accuracy of a model, it model needs to be fit to the data.

The current results suggest that while discounting models can account for many behavioral effects seen in risky intertemporal choice, they are a poor statistical fit to choices. Given the predictive power and process evidence of heuristic models in *independent* risky and intertemporal choices, it follows that heuristics operate in *combined* risky intertemporal choices. Recent work on risky intertemporal choices with only two outcomes, found that people spend relatively more time fixating on risk compared to time (Konstantinidis, van Ravenzwaaij, & Newell, 2017). When there are more probabilities—e.g. there is amount risk—people may focus even more on probability, decreasing their relative focus on the delay—i.e., if people don't pay attention to time then they should seem patient. Decreased relative attention means that delay plays a lesser role in choice, which could explain why amount risk has a bigger effect on discounting than outcome risk.

Konstantinidis, et al. also showed that information acquisition is inconsistent with amount and time both falling under psychological distance. We add to their result by showing that while risk and time are entangled choice models are inconsistent with risk and time both falling under psychological distance. Our results also suggest that a functional form for the discount function of risky intertemporal choices is difficult to detect and that new models are needed.

Acknowledgements

We thank Ashwani Monga, Daniel Oppenhiemer, and Stephen Broomell for their comments.

References

Andreoni, J., & Sprenger, C. (2012). Risk Preferences Are Not Time Preferences. *American Economic Review*,

Baucells, M., & Heukamp, F. H. (2012). Probability and Time Trade-Off. *Management Science*, *58*(4), 831–842.

Blackburn, M., & El-Deredy, W. (2013). The future is risky: Discounting of delayed and uncertain outcomes. *Behavioural Processes*, *94*, 9–18.

Broomell, S. B., & Bhatia, S. (2014). Parameter Recovery for Decision Modeling Using Choice Data. *Decision*, *1*(4), 252–274.

Ebert, J. E. J., & Prelec, D. (2007). The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future. *Management Science*, *53*(9), 1423–1438.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzod, S., Hau, R., … Lebiere, C. (2014). A Choice Prediction Competition: Choices from Experience and from Description. *The Journal of Behavioral Decision Making*, *27*, 179–189.

Gonzalez, R., & Wu, G. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, *38*, 129–166.

Hardisty, D. J., & Pfeffer, J. (2015). Intertemporal Uncertainty Avoidance: When the Future is Uncertain, People Prefer the Present, and When the Present is Uncertain, People Prefer the Future. *Management Science*, *53*(November), 160.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(March), 263–291.

Knoll, M. A., Appelt, K. C., Johnson, E. J., & Westfall, J. E. (2015). Time to Retire: Why Americans Claim Benefits Early & how to encourage delay. *Behavioral Science & Policy*, *1*(1), 53–62.

Konstantinidis, E., van Ravenzwaaij, D., & Newell, B. R. (2017). Exploring the decision dynamics of risky intertemporal choice. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 694–699.

Nilsson, H., Rieskamp, J., & Wagenmakers, E. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*(1), 84–93.

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497–527.

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*,

Schreiber, P., & Weber, M. (2016). The Influence of Time Preferences on Retirement Timing, (May), 1–2.

Toubia, O., Johnson, E., Evgeniou, T., & Delquié, P. (2012). Dynamic Experiments for Estimating Preferences : An Adaptive Method of Eliciting Time and Risk Parameters. *Management Science*, *1909*, 1–28.

Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance Yaacov. *Psychological Review2*, *117*(2), 440–463.

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

Weber, B. J., & Chapman, G. B. (2005). The combined effects of risk and time on choice: Does uncertainty eliminate the immediacy effect? Does delay eliminate the certainty effect? *Organizational Behavior and Human Decision Processes*, *96*(2), 104–118.